BOLYAI SOCIETY
MATHEMATICAL STUDIES

20

Gyula O. H. Katona
Alexander Schrijver
Tamás Szőnyi
Editors

Fete of Combinatorics and Computer Science

Springer

JBMS

BOLYAI SOCIETY
MATHEMATICAL STUDIES 20

BOLYAI SOCIETY MATHEMATICAL STUDIES

1. **Combinatorics, Paul Erdős is Eighty, Vol. 1**
   D. Miklós, V.T. Sós, T. Szőnyi (Eds.)
2. **Combinatorics, Paul Erdős is Eighty, Vol. 2**
   D. Miklós, V.T. Sós, T. Szőnyi (Eds.)
3. **Extremal Problems for Finite Sets**
   P. Frankl, Z. Füredi, G. Katona, D. Miklós (Eds.)
4. **Topology with Applications**
   A. Császár (Ed.)
5. **Approximation Theory and Function Series**
   P. Vértesi, L. Leindler, Sz. Révész, J. Szabados, V. Totik (Eds.)
6. **Intuitive Geometry**
   I. Bárány, K. Böröczky (Eds.)
7. **Graph Theory and Combinatorial Biology**
   L. Lovász, A. Gyárfás, G. Katona, A. Recski (Eds.)
8. **Low Dimensional Topology**
   K. Böröczky, Jr., W. Neumann, A. Stipsicz (Eds.)
9. **Random Walks**
   P. Révész, B. Tóth (Eds.)
10. **Contemporary Combinatorics**
    B. Bollobás (Ed.)
11. **Paul Erdős and His Mathematics I+II**
    G. Halász, L. Lovász, M. Simonovits, V. T. Sós (Eds.)
12. **Higher Dimensional Varieties and Rational Points**
    K. Böröczky, Jr., J. Kollár, T. Szamuely (Eds.)
13. **Surgery on Contact 3-Manifolds and Stein Surfaces**
    B. Ozbagci, A. I. Stipsicz
14. **A Panorama of Hungarian Mathematics in the Twentieth Century, Vol. 1**
    J. Horváth (Ed.)
15. **More Sets, Graphs and Numbers**
    E. Győri, G. Katona, L. Lovász (Eds.)
16. **Entropy, Search, Complexity**
    I. Csiszár, G. Katona, G. Tardos (Eds.)
17. **Horizons of Combinatorics**
    E. Győri, G. Katona, L. Lovász (Eds.)
18. **Handbook of Large-Scale Random Networks**
    B. Bollobás, R. Kozma, D. Miklós (Eds.)
19. **Building Bridges**
    M. Grötschel, G. Katona (Eds.)

Gyula O. H. Katona
Alexander Schrijver
Tamás Szőnyi
(Eds.)

Fete of Combinatorics and Computer Science

Springer
JÁNOS BOLYAI MATHEMATICAL SOCIETY

Gyula O. H. Katona
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
Reáltanoda u. 13-15
Budapest 1053
Hungary
e-mail: ohkatona@renyi.hu

Alexander Schrijver
Centre for Mathematics
and Computer Science
Science Park 123
1098 XG Amsterdam
The Netherlands
e-mail: lex.schrijver@cwi.nl

Tamás Szőnyi
Eötvös University
Department of Computer Science
Pázmány Péter sétány 1/C
Budapest 1117
Hungary
e-mail: szonyi@cs.elte.hu

Managing Editor:
Gábor Sági
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
Reáltanoda u. 13-15
Budapest 1053
Hungary
e-mail: sagi@renyi.hu

Mathematics Subject Classification (2000): 05XX, 68XX

Library of Congress Control Number: 2010928477

ISSN 1217-4696
ISBN 978-3-642-13579-8 Springer Berlin Heidelberg New York
ISBN 978-963-9453-12-8 János Bolyai Mathematical Society, Budapest

Printed in Hungary

Cover design: WMXDesign GmbH, Heidelberg

Printed on acid-free paper 44/3142/db – 5 4 3 2 1 0

Contents

PREFACE

This book is the second volume of mathematical papers to commemorate the 60th birthday of László ('Laci') Lovász on March 9, 2008. Prominent mathematicians that were inspired by Laci were invited to contribute to the celebrations of this milestone in the Summer of 2008. The response was overwhelming, and therefore two conferences were organized, the first from 5–9 August 2008 in Budapest, and the second the week after, from 11–15 August 2008 in Keszthely, at the borders of Lake Balaton. The first meeting was baptized *Building Bridges – Between Mathematics and Computer Science,* the second *Fete of Combinatorics and Computer Science.* These names also adorn the two volumes published to commemorate the celebrations.

The two names very well reflect the power and the beauty of Laci's work. Many of his results indeed can be characterized by building new, powerful bridges between and within mathematics and computer science, and these are really feasts to the combinatorial and algorithmic mind. Several of the methods found by Lovász have opened up new areas of research in discrete mathematics and algorithmics. Often, these methods were obtained by Lovász by developing beautiful new techniques based on algebra, geometry, or topology, thus laying fundaments for the important role of classical mathematics in more modern fields like combinatorics, optimization, algorithmics, and complexity.

Laci Lovász has obtained so many pioneering results that it would be quite impossible to describe a reasonable part of it. Let us just give a brief selection. He solved several important open problems in discrete mathematics, like the perfect graph conjecture, Kneser's conjecture, the Shannon capacity problem, the matroid matching problem, and the matching lattice problem. He proved the now called 'Lovász Local Lemma', basic in probabilistic combinatorics and randomized algorithms. He pointed out the significance of the ellipsoid method and semidefinite programming for combinatorial optimization, and he designed the lattice basis reduction method, with a wealth of applications in discrete mathematics, combinatorial optimization, integer programming, number theory, algebra, and cryptography. Other basic algorithmic results were found for volume computation and

communication complexity. Recently, Laci was the main inspirator of the new area of graph limits, with its connections to extremal combinatorics and mathematical physics.

Laci's deep insight in the combinatorial and algorithmic potentials of geometry and topology is unique. Several of his results are ground-breaking, and there seems little chance that they would have been detected without him. They have inspired and directed the work of a huge part of the researchers in discrete mathematics, optimization, and algorithmics.

Laci's reputation is beyond any measure not only because of his results and insights, but also because of his very pleasant modesty, his willingness to cooperate and to exchange ideas, and the extreme transparency of his lectures, which form usually exciting eye-openers.

The inspiration and charismatic ambiance offered by Laci and his wife Kati have given them many friends and collaborators, all over the world and in a wide range in mathematics and computer science. We are very happy that many of them have reacted enthusiastically to our invitation to contribute to the volumes, and that we received so many excellent articles. Like the first volume, also this second volume forms an expression of the impact of Laci Lovász's work and of the inspiration given by him to many scientists world-wide.

Budapest
August 2009

Gyula Katona
Alexander Schrijver
Tamás Szőnyi

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 9–21.

HIGH DEGREE GRAPHS CONTAIN LARGE-STAR FACTORS

NOGA ALON* and NICHOLAS WORMALD†

Dedicated to László Lovász, for his 60th birthday

We show that any finite simple graph with minimum degree d contains a spanning star forest in which every connected component is of size at least $\Omega((d/\log d)^{1/3})$. This settles a problem of Havet, Klazar, Kratochvil, Kratsch and Liedloff.

Dedication

This paper is dedicated to Laci Lovász, for his 60th birthday. It settles a problem presented by Jan Kratochvil at the open problems session of the meeting *Building Bridges,* which took place in Budapest in August 2008, celebrating this birthday. The Lovász Local Lemma is applied extensively throughout the proof. This work is therefore a typical example illustrating the immense influence of Laci, who not only provided the community with powerful tools and techniques, but also stimulated research by his books, lectures and organization of conferences.

1. INTRODUCTION

All graphs considered here are finite and simple. A *star* is a tree with one vertex, the *center,* adjacent to all the others, which are *leaves.* A *star factor* of a graph G is a spanning forest of G in which every connected component is a star. It is easy to see that any graph with positive minimum degree contains a star factor in which every component is a star with at least one

*Research supported in part by ERC advanced grant, by a USA-Israel BSF grant, by NSF grant CCF 0832797 and by the Ambrose Monell Foundation.

†Supported by the Canada Research Chairs Program and NSERC.

edge. A conjecture of Havet et al. [9], communicated to us by Jan Kratochvil [10], asserts that if the minimum degree is large then one can ensure that all stars are large. More precisely, they conjectured that there is a function $g(d)$ that tends to infinity as d tends to infinity, so that every graph with minimum degree d contains a star factor in which every star contains at least $g(d)$ edges. Our main result shows that this is indeed the case, for a function $g(d)$ that grows moderately quickly with d, as follows.

Theorem 1.1. *There exists an absolute positive constant c so that for all $d \geq 2$, every graph with minimum degree d contains a star factor in which every star has at least $cd^{1/3}/(\log d)^{1/3}$ edges.*

The motivation for the conjecture of Havet et al. arises in the running time analysis of a recent exact exponential time algorithm for the so called $L(2,1)$-labeling problem of graphs. See [9] for more details.

As preparation for the proof of the main result, we prove the following simpler statement.

Theorem 1.2. *There exists an absolute positive constant c' such that for all $d \geq 2$, every d-regular graph contains a star factor in which every star has at least $c'd/\log d$ edges. This is optimal, up to the value of the constant c'.*

Throughout the paper we make no attempt to optimize the absolute constants. To simplify the presentation we omit all floor and ceiling signs whenever these are not crucial. We may and will assume, whenever this is needed, that the minimum degree d considered is sufficiently large. It is easy to find, in any graph with all vertices of degree at least 1, a star factor with stars of at least two vertices each, so the theorems then follow for all $d \geq 2$. All logarithms are in the natural base, unless otherwise specified.

Our notation is standard. In particular, for a graph $G = (V, E)$ and a vertex $v \in V$, we let $N_G(v)$ denote the set of all neighbors of v in the graph G, and let $d_G(v) = \left|N_G(v)\right|$ denote the degree of v in G. For $X \subset V$, $N_G(X) = \cup_{x \in X} N_G(x)$ is the set of all neighbors of the members of X.

The rest of this short paper is organized as follows. In Section 2 we present the simple proof of Theorem 1.2, and in Section 3 the proof of the main result. Section 4 contains some concluding remarks and open problems.

2. Regular Graphs

Proof of Theorem 1.2. Let $G = (V, E)$ be a d-regular graph. Put $p = (2 + 2\log d)/d$ and let C be a random set of vertices obtained by picking each vertex of G, randomly and independently, to be a member of C, with probability p. We will show that, with positive probability, some such set C will be a suitable choice for the set of centres of the stars in the desired star factor.

For each vertex $v \in V$, let A_v be the event that either v has no neighbors in C or v has more than $3pd$ neighbors in C. By the standard known estimates for binomial distributions (c.f., e.g., [4], Theorem A.1.12), the probability of each event A_v is at most $(1-p)^d + (e^2/27)^{2+2\log d} < e^{-p} + (1/e)^{2+2\log d} < 1/ed^2$. Moreover, each event A_v is mutually independent of all events A_u except those that satisfy $N_G(v) \cap N_G(u) \neq \emptyset$. As there are at most $d(d-1) < d^2$ such vertices u we can apply the Lovász Local Lemma (c.f., e.g., [4], Corollary 5.1.2) to conclude that with positive probability none of the events A_v holds. Therefore, there is a choice of a set $C \subset V$ so that for every vertex v, $0 < \left|N_G(v) \cap C\right| \leq 3pd = 6 + 6\log d$.

Fix such a set C, and let B be the bipartite graph whose two classes of vertices are C and $V \setminus C$, where each $v \in C$ is adjacent in B to all vertices $u \in V \setminus C$ which are its neighbors in G. By the choice of C, for every vertex $v \in C$, $d_B(v) \geq d - 6 - 6\log d$ and for every vertex $u \in V \setminus C$, $0 < d_B(u) \leq 6 + 6\log d$. It thus follows by Hall's theorem that one can assign to each vertex $v \in C$ a set consisting of

$$\frac{d - 6 - 6\log d}{6 + 6\log d} > \frac{d}{7\log d}$$

of its neighbors in $V \setminus C$, where no member of $V \setminus C$ is assigned to more than one such v. (To see this from the standard version of Hall's theorem, split each vertex v in C into $(d-6-6\log d)/(6+6\log d)$ identical 'sub-vertices', each with the same neighbours in $V \setminus C$ as v, and find a matching that hits every sub-vertex using Hall's theorem. Then for each $v \in C$, coalesce the subvertices of v back together to form v.) By assigning each unassigned vertex u of $V \setminus C$ arbitrarily to one of its neighbors in C (note that there always is such a neighbor) we get the required star factor, in which the centers are precisely the members of C and each star contains more than $d/(7\log d)$ edges.

It remains to show that the above estimate is optimal, up to a constant factor. Note that the centers of any star factor form a dominating set in the graph, and thus if the minimum size of a dominating set in a d-regular graph on n vertices is at least $\Omega(n \log d/d)$, then the star factor must contain a component of size at most $O(d/\log d)$. It is not difficult to check that the minimum size of a dominating set in a random d-regular graph on n vertices is $\Theta(n \log d/d)$ with high probability. In fact, if $c < 1$, the expected number of dominating sets of size $k = n\big(c + o(1)\big)(\log d)/d$ tends to 0 for d sufficiently large. We give some details of verifying this claim. One can use the standard pairing or configuration model of random d-regular graphs, in which there are n buckets with d points in each bucket. It is enough to prove the result for the multigraph arising from taking a random pairing of the points and regarding the buckets as vertices (see e.g. [12] for details). The expected number of dominating sets S of size k with m edges from S to $N(S)$ is

$$A := \binom{n}{k} f(k,m) \Big(\prod_{i=0}^{m-1} (kd - i) \Big) \frac{M(kd-m)M\big((n-k)d-m\big)}{M(nd)}$$

where $f(k,m)$ is the coefficient of x^m in $\big((1+x)^d - 1\big)^{n-k}$ and $M(r)$ is the number of pairings or perfect matchings of an even number r of points, i.e. $(r-1)(r-3)\cdots 1$. In the above formula, the binomial chooses the k buckets of S, $f(k,m)$ is the number of ways to choose m points in the other $n-k$ buckets such that at least one point comes from each bucket (so that S dominates the graph), and the next factor counts the ways to pair those points with points in buckets in S. The other factors in the numerator count the ways to pair up the remaining points, first within S, and then within the rest of the graph. The denominator is the total number of pairings. We may use standard methods to see that $f(k,m) \le \big((1+x)^d - 1\big)^{n-k} x^{-m}$ for all real $x > 0$ (since f's coefficients are all nonnegative). We set $x = k/n$ and take $k = nc(\log d)/d$, which is justified by regarding c as a function of n that tends to a limit equal to the value c referred to in the claim above. This gives

$$A \le \binom{n}{k} \big((1+k/n)^d - 1\big)^{n-k} \Big(\frac{n}{k}\Big)^m \frac{(kd)!}{(kd-m)!} \cdot \frac{M(kd-m)M\big((n-k)d-m\big)}{M(nd)}.$$

Considering replacing m by $m+2$ shows this expression is maximised when $n^2(kd-m) \approx k^2\big((n-k)d-m\big)$ and certainly only when $kd-m \sim n\big((c\log d)^2/d\big)$. In fact, $kd-m = O\big(n(\log^2 d)/d\big)$ is sufficient for our purposes. Fixing such a value of m, using $n! = (n/e)^n n^{\theta(1)}$ and $M(r) = \Theta\big((r/e)^{r/2}\big)$, and noting that

$$\binom{n}{k} = \exp\big(O\big(n(\log^2 d)/d\big)\big),$$

$$(1+k/n)^d - 1 = d^c\big(1 - d^{-c} + O\big(d^{-1}\log^2 d\big)\big),$$

$$\left(\frac{n}{k}\right)^m = \left(\frac{n}{k}\right)^{-(kd-m)}\left(\frac{n}{k}\right)^{kd} = \left(\frac{n}{k}\right)^{kd}\exp\big(O\big(n(\log^3 d)/d\big)\big),$$

$$\big((kd-m)/e\big)^{kd-m} = n^{kd-m}\exp\big(O\big(n(\log^3 d)/d\big)\big),$$

$$\big((n-k)d-m\big)^{((n-k)d-m)/2} = \big(nd-2kd+(kd-m)\big)^{((n-k)d-m)/2}$$

$$= (nd)^{((n-k)d-m)/2}\left(1 - \frac{2k}{n} + \frac{k-m/d}{2n}\right)^{nd\left(1-2k/n+(k-m/d)/2n\right)/2}$$

$$= (nd)^{((n-k)d-m)/2}\exp\big(-dk + O\big(n(\log^2 d)/d\big)\big)$$

we find everything in the upper bound for A cancels except for

$$e^{-kd}d^{c(n-k)+(kd-m)/2}(1-d^{-c})^n\exp\big(O\big(nd^{-1}\log^3 d\big)\big).$$

Since $d^c = e^{kd/n}$ and the power of d is absorbed in the error term, this equals $n^{O(1)}\big(1-d^{-c}+O(d^{-1}\log^3 d)\big)^n$, which, if $c<1$, tends to 0 for large d as required. On the other hand, for $c>1$ and large d, $nc(\log d)/d$ is an upper bound on the minimum dominating set size in all d-regular graphs [4, Theorem 2.2].

An explicit example can be given as well: if $d = (p-1)/2$ with p being a prime, consider the bipartite graph H with two classes of vertices $A_1 = A_2 = Z_p$ in which $a_i b_j$ forms an edge iff $(a_i - b_j)$ is a quadratic non-residue. A simple consequence of Weil's Theorem (see, e.g., [2], Section 4) implies that for every set S of at most, say, $\frac{1}{3}\log_2 p$ elements of Z_p, there are more than $\sqrt{p}$ members z of Z_p so that $(z-s)$ is a quadratic residue for all $s \in S$. This implies that any dominating set of H must contain either more than $\frac{1}{3}\log_2 p$ vertices of A_2 or at least $\sqrt{p}$ vertices of A_1, and is thus of

size bigger than $\frac{1}{3}\log_2 p$. (This can in fact be improved to $\big(1-o(1)\big)\log_2 p$, but as we are not interested in optimizing the absolute constants here and in the rest of the paper, we omit the proof of this stronger statement). For degrees d that are not of the form $(p-1)/2$ for a prime p one can take any spanning d-regular subgraph of the graph above with the smallest p for which $(p-1)/2 \geq d$. Such a subgraph exists by Hall's theorem, and any dominating set in it is also dominating in the original $(p-1)/2$-regular graph, hence it is of size at least $\frac{1}{3}\log_2 p$. By the known results about the distribution of primes this prime p is $\big(2+o(1)\big)\,d$, and we thus get a d-regular graph on at most $n = \big(4+o(1)\big)\,d$ vertices in which every dominating set is of size greater than $\frac{1}{3}\log_2 p > \frac{1}{3}\log_2 d \geq \Omega\big(n(\log d)/d\big)$. This completes the proof. ■

3. The Proof of the Main Result

In this section we prove Theorem 1.1. The idea of the proof is based on that of Theorem 1.2. Given a graph $G = (V, E)$ with minimum degree d, we wish to define a dominating set $C \subset V$ whose members will form the centers of the star factor, and then to assign many leaves to each of them. The trouble is that here we cannot pick the set of centers randomly, as our graph may contain a large set R of vertices of degree d whose total number of neighbors is much smaller than $|R|$, and then the number of centers in R is limited. This may happen if some or all of the neighbors of the vertices in R have degrees which are much higher than d. Thus, for example, if our graph is a complete bipartite graph with classes of vertices R and U, with $|R| = n - d$ and $|U| = d \ll n - d$, it is better not to choose any centers in R. In fact, it seems reasonable in the general case to force all vertices of degree much higher than d to be centers, and indeed this is the way the proof starts. However, if we then have a vertex all (or almost all) of whose neighbors have already been declared to be centers, then this vertex cannot be a center itself, and will have to be a leaf. Similarly, if almost all neighbors of a vertex are already declared to be leaves, then this vertex will have to become a center.

The proof thus proceeds by declaring, iteratively, some vertices to be centers and other vertices to be leaves. At the end, if there are any vertices left, we choose a small subset of them randomly to be additional centers.

The Local Lemma has to be applied to maintain the desired properties that will enable us to apply Hall's theorem at the end to a bipartite graph, defined in a way similar to that in the proof of Theorem 1.2. An additional complication arises from the fact that we have to assign time labels to vertices and use them in the definition of the bipartite graph. We proceed with the detailed proof.

Proof of Theorem 1.1. Let $G = (V, E)$ be a graph with minimum degree d. We first modify G by omitting any edge whose two endpoints are of degree strictly greater than d, as long as there is such an edge. We thus may and will assume, without loss of generality, that every edge has at least one endpoint of degree exactly d. Put $h = \frac{1}{10}d^{4/3}/(\log d)^{1/3}$, and let H (for *High*) denote the set of all vertices of degree at least h. Since each of their neighbors is of degree precisely d, we can apply Hall's theorem and assign a set of h/d neighbors to each of them, so that no vertex is assigned twice. Let S' denote the set of all the $|H| \cdot \frac{1}{10}d^{1/3}/(\log d)^{1/3}$ assigned vertices, and let S (for *Special*) be a random subset of S' obtained by choosing each member of S' to be in S randomly and independently with probability $1/2$.

Claim 3.1. *In the random choice of S, with positive probability the following conditions hold:*
(i) *For each $v \in H$, $\left|N_G(v) \cap S\right| > \frac{1}{25}d^{1/3}/(\log d)^{1/3}$.*
(ii) *For each $v \in V \setminus H$, $\left|N_G(v) \setminus S\right| \geq d/3$.*

Proof. For each vertex $v \in H$ let A_v be the event that condition (i) is violated for v. Similarly, for each vertex $v \in V \setminus H$ let B_v be the event that condition (ii) is violated for v. By the standard known estimates for binomial distributions, the probability of each event A_v is $\exp\left(-\Omega\left(d^{1/3}/(\log d)^{1/3}\right)\right)$ and that of each event B_v is $e^{-\Omega(d)}$. In addition, each event A_v is independent of all other events except for the events B_u for vertices u that have a neighbor among the $\frac{1}{10}d^{1/3}/(\log d)^{1/3}$ vertices of S' assigned to v (note that there are less than $d^{4/3}$ such vertices u). The same reasoning shows that each event B_v is mutually independent of all other events A_u, B_w with the exception of at most $hd < d^3$ events. The desired result thus follows from the Local Lemma (with a lot of room to spare). This completes the proof of the claim.

Fix an S satisfying the assertion of the claim, and define $G' = G - S = (V', E')$. Note that by the above claim, part (ii),

$$\text{for each} \quad v \in V(G') = V \setminus S, \quad \left|N_{G'}(v)\right| \geq d/3. \tag{1}$$

Note also that by part (i) of the claim, each vertex $v \in H$ has a set of at least $\frac{1}{25}d^{1/3}/(\log d)^{1/3}$ vertices from S assigned to it, and can thus serve as a center of a star of at least that size.

We now construct two sets of vertices $C, L \subset V \setminus S$. The set C will consist of vertices that are declared to be *centers*, and will serve as centers of stars in our final star factor. The set L will consist of vertices that are declared to be *leaves* in the final factor. Note that the vertices in S will not form part of these sets; they will also be leaves in the final star factor, and their associated centers will be the vertices in H to which they have been assigned, but since we have already specified their centers we do not need to consider them any more. Initially, define $C = H$ and $L = \emptyset$. We will also need a time label $t(v)$ which will be defined in the following for each vertex in $V \setminus (H \cup S)$; in the beginning set $t = 0$.

Put $V' = V(G') = V \setminus S$ and define $D = d^{2/3}(\log d)^{1/3}$. Now repeatedly apply the following two rules to define additional centers and leaves, and assign them time labels.

- **(a)** If there is a vertex $v \in V' \setminus (C \cup L)$ such that $\left|N_{G'}(v) \setminus C\right| \leq D$, add v to L, increase t by 1, and define $t(v) = t$.

- **(b)** If there is a vertex $v \in V' \setminus (C \cup L)$ such that $\left|N_{G'}(v) \setminus L\right| \leq d/6$, add v to C, increase t by 1, and define $t(v) = t$.

The process continues by repeatedly applying rules (a) and (b) in any order until there are no vertices left in $V' \setminus (C \cup L)$ that satisfy the conditions in rule (a) or in rule (b). Let t_0 denote the value of the time parameter t at this point. Actually, since L and C will be disjoint, by (1) no vertex will satisfy the conditions in both rules simultaneously. Let F (for *Free*) denote the set of all vertices in $V' \setminus (C \cup L)$ remaining once the process terminates. Define $p = 20(\log d)/d$, and let T be a random subset of F obtained by picking each vertex $v \in F$, randomly and independently, to be in T with probability p. Assign the vertices of $F \setminus T$ the time labels $t_0 + 1, t_0 + 2, \ldots, t_0 + |F \setminus T|$ in any order. Finally, assign the vertices of T the time labels $t_0 + |F \setminus T| + 1, t_0 + |F \setminus T| + 2, \ldots, t_0 + |F|$.

In our final star factor, the vertices $H \cup C \cup T$ will serve as centers, while the remaining vertices, that is, those in $S \cup L \cup (F \setminus T)$, will serve as leaves. In order to show that it is possible to define large stars with these centers and leaves, we need the following.

Claim 3.2. *With positive probability, every vertex of F has at least one neighbor in $C \cup T$, and no vertex $v \in V' - H$ has more than $2ph = 4d^{1/3}(\log d)^{2/3}$ neighbors in T.*

Proof. For each vertex $v \in F$ that does not have any neighbor in C, let A_v be the event that it has no neighbor in T. Note that as $v \in F$, the definition of rule (b) implies that $\left|N_{G'}(v) \setminus L\right| > d/6$, and as it has no neighbor in C, it has more than $d/6$ neighbors in F. Therefore, the probability that none of these neighbors is in T is at most $(1-p)^{d/6} < d^{-3}$. For each vertex $v \in V' \setminus H$, let B_v be the event that v has more than $2ph$ neighbors in T. Since the degree of v in G' is at most h, its number of neighbors in F is certainly at most h, and hence the standard estimates for binomial distributions imply that the probability of each event B_v is at most $e^{-\Omega(ph)}$ which is much smaller than, say, d^{-3}.

Note that each event A_v is mutually independent of all other events A_u or B_w apart from those corresponding to vertices u or w that have a common neighbor with v in F, and the number of such vertices u, w is smaller than $hd < d^{7/3}$. Similarly, each of the events B_v is independent of all others but at most $hd < d^{7/3}$. The claim thus follows from the Local Lemma.

Returning to the proof of the theorem, fix a choice of F satisfying the assumptions in the last claim. Let B be the bipartite graph with classes of vertices $(C \setminus H) \cup T$ and $L \cup (F \setminus T)$, in which each $v \in (C \setminus H) \cup T$ is adjacent to any of its neighbors u that lies in $L \cup (F \setminus T)$ and satisfies $t(u) < t(v)$. Note that, crucially, prospective centers are connected in B only to prospective leaves with smaller time labels.

Our objective is to show, using Hall's theorem, that we can assign to each vertex v in $(C \setminus H) \cup T$ some $\Omega(d^{1/3}/(\log d)^{1/3})$ neighbors of v (in B, and hence also in G') from $L \cup (F \setminus T)$, such that each vertex in $L \cup (F \setminus T)$ is assigned at most once. To do so, we first establish several simple properties of the bipartite graph B that follow from its construction.

Claim 3.3. *The following properties hold.*

(i) *For each vertex $u \in L$, $\left|N_B(u) \cap (C \setminus H)\right| \leq D = d^{2/3}(\log d)^{1/3}$.*

(ii) *For each vertex $u \in L$, $\left|N_B(u) \cap T\right| \leq 4d^{1/3}(\log d)^{2/3}$.*

(iii) *For each vertex $u \in F \setminus T$, $\left|N_B(u)\right| = \left|N_B(u) \cap T\right| \leq 4d^{1/3}(\log d)^{2/3}$.*

(iv) *For each vertex $v \in C \setminus H$, $d_B(v) \geq d/6$.*

(v) *For each vertex $v \in T$, $\left|N_B(v)\right| \geq D - 4d^{1/3}(\log d)^{2/3} > D/2 = \frac{1}{2}d^{2/3}(\log d)^{1/3}$.*

Proof.

(i) By the definition of rule (a), each $u \in L$ can have at most D neighbors with time labels exceeding $t(u)$, and therefore can have at most that many neighbors in B.

(ii) This follows immediately from the condition in Claim 3.2 that F was chosen to satisfy.

(iii) By the definition of the graph B, the vertices in $F \setminus T$ are joined in B only to vertices of T, as these are the only vertices with bigger time labels. Therefore, $\left|N_B(u)\right| = \left|N_B(u) \cap T\right|$ for each $u \in F \setminus T$, and the claimed upper estimate for this cardinality follows from Claim 3.2.

(iv) By the definition of rule (b), each vertex $v \in C \setminus H$ satisfied $\left|N_{G'}(v) \setminus L\right| \leq d/6$ at the point of being added to C. Since by (1), $\left|N_{G'}(v)\right| \geq d/3$, it follows that at that time, v had at least $d/3 - d/6 = d/6$ neighbors in L. As all these leaves have smaller time labels than v, it is joined in B to all of them.

(v) If $v \in T$, then $v \in F$, and thus, by the definition of rule (a), $\left|N_{G'}(v) \setminus C\right| > D$. Vertices in T are given the largest time labels, so in the graph B, v is joined to all members of $N_{G'}(v) \setminus C$ except for those that lie in T. However, by the condition in Claim 3.2, at most $4d^{1/3}(\log d)^{2/3}$ of these vertices are members of T, implying the desired estimate. This completes the proof of the claim.

Corollary 3.1. *For each subset* $X \subset (C \setminus H) \cup T$, $\left|N_B(X)\right| \geq |X| \cdot \frac{1}{16}d^{1/3}/(\log d)^{1/3}$.

Proof. If at least half the elements of X belong to $C \setminus H$, then, by Claim 3.3(iv), the total number of edges of B incident with them is at least $\frac{1}{2}|X| \cdot \frac{1}{6}d$. By the first observation in the proof of part (iii) of the claim, these edges are not incident in B with any member of $F \setminus T$. By part (i) of the claim, at most $D = d^{2/3}(\log d)^{1/3}$ of these edges are incident with any one vertex in L. Thus, in this case, $\left|N(X)\right| \geq \frac{1}{2}|X| \cdot \frac{1}{6}d \cdot 1/D = |X| \cdot \frac{1}{16}d^{1/3}/(\log d)^{1/3}$, providing the required estimate.

Otherwise, at least half of the vertices of X lie in T. By Claim 3.3, part (v), the total number of edges of B incident with them is greater than $\frac{1}{2}|X| \cdot \frac{1}{2}D$. By parts (ii) and (iii) of the claim, each neighbor of these vertices in $L \cup (F \setminus T)$ is incident with at most $4d^{1/3}(\log d)^{2/3}$ of these edges, implying

that in this case

$$|N(X)| \geq \frac{|X|}{2}\frac{D}{2}\frac{1}{4d^{1/3}(\log d)^{2/3}} = \frac{d^{1/3}}{16(\log d)^{1/3}}|X|.$$

This completes the proof of the corollary.

By the last Corollary and Hall's theorem, one can assign a set of $\frac{1}{16}d^{1/3}/(\log d)^{1/3}$ members of $L \cup (F \setminus T)$ to any element of $(C \setminus H) \cup T$, so that no member of $L \cup (F \setminus T)$ is assigned more than once. This makes all elements of $(C \setminus H) \cup T$ centers of large vertex disjoint stars. Adding to these stars the stars whose centers are the elements of H and whose leaves are those of S, we may apply Claim 3.2 to conclude that if there are any unassigned vertices left in L we can connect each of them to one of the existing centers. Similarly, (1) and the definition of rule (a) do the same job for unassigned vertices in $F \setminus T$. Thus, we get a star factor in which each star has at least $\frac{1}{25}d^{1/3}/(\log d)^{1/3}$ leaves. This completes the proof. ■

4. Concluding Remarks and Open Problems

We have shown that for every positive integer g there is an integer d so that any graph with minimum degree at least d contains a star factor in which every component has at least g edges. Let $d(g)$ denote the minimum number d for which this holds. Our main result shows that $d(g) \leq O(g^3 \log g)$, while the construction described in the proof of Theorem 1.2 implies that $d(g) \geq \Omega(g \log g)$. It seems plausible to conjecture that $d(g) = \Theta(g \log g)$, but this remains open. It will be interesting to determine $d(g)$ precisely (or estimate it more accurately) for small values of g, like $g = 2$ or 3. Our proof, even if we try to optimize the constants in it, will yield only some crude upper bounds that are certainly far from being tight. It is worth noting, however, that even for showing that $d(2)$ is finite, we do not know any proof simpler than the one given here for the general case. On the other hand, a random 3-regular graph contains a Hamilton cycle with probability tending to 1. Hence, if it has number of vertices divisible by 3, it contains a spanning factor of stars of two edges each. Similarly, a random 4-regular graph having number of vertices divisible by 4 contains a spanning factor of stars of three edges each with probability tending to 1 [5]. Immediately

from contiguity results discussed in [12], the same statements are true if we change 4-regular to d-regular for any $d \geq 4$.

Our proof, together with the algorithmic version of the local lemma proved by Beck in [6] (see also [1]), and any efficient algorithm for bipartite matching, show that the proof here can be converted to a deterministic, polynomial time algorithm that finds, in any given input graph with minimum degree at least d, a star factor in which every star is of size at least $\Omega\big((d/\log d)^{1/3}\big)$. We omit the details.

There are several known results that show that any connected graph with large minimum degree contains a spanning tree with many leaves, see [11], [8], [7]. In particular, it is known (and not difficult) that any connected graph with minimum degree d and n vertices contains a spanning tree with at least $n - O\big(n(\log d)/d\big)$ leaves. A related question to the one considered here is whether it is true that any connected graph with large minimum degree contains a spanning tree in which all non-leaf vertices have large degrees. Specifically, is there an absolute positive constant c so that any connected graph with minimum degree at least d contains a spanning tree in which the degree of any non-leaf is at least $cd/\log d$? Another intriguing question is the following possible extension of the main result here. Is it true that the edges of any graph G with minimum degree d can be partitioned into pairwise disjoint sets, so that each set forms a spanning star forest of G in which every component is of size at least $h(d)$, where $h(d)$ tends to infinity with d? A related result is proved in [3], but the proof of the last statement, if true, seems to require additional ideas.

References

[1] N. Alon, A parallel algorithmic version of the local lemma, *Random Structures and Algorithms,* **2** (1991), 367–378.

[2] N. Alon, Tools from higher algebra, in: *Handbook of Combinatorics* (R. L. Graham, M. Grötschel and L. Lovász, eds), North Holland (1995), Chapter 32, pp. 1749–1783.

[3] N. Alon, C. McDiarmid and B. Reed, Star arboricity, *Combinatorica,* **12** (1992), 375–380.

[4] N. Alon and J. H. Spencer, *The Probabilistic Method,* Third Edition, Wiley (New York, 2008).

[5] H. Assiyatun and N. Wormald, 3-star factors in random regular graphs, *Europ. J. Combinatorics,* **27** (2006), 1249–1262.

[6] J. Beck, An algorithmic approach to the Lovász Local Lemma, *Random Structures and Algorithms,* **2** (1991), 343–365.

[7] Y. Caro, D. B. West and R. Yuster, Connected domination and spanning trees with many leaves, *SIAM J. Discrete Math.,* **13** (2000), 202–211.

[8] J. R. Griggs and M. Wu, Spanning trees in graphs of minimum degree four or five, Discrete *Mathematics,* **104** (1992), 167–183.

[9] F. Havet, M. Klazar, J. Kratochvil, D. Kratsch and M. Liedloff, Exact algorithms for $L(2,1)$-labelling, *Algorithmica,* to appear.

[10] J. Kratochvil, An open problem presented in "Building Bridges", Budapest, 2008.

[11] D. J. Kleitman and D. B. West, Spanning trees with many leaves, *SIAM Journal on Discrete Mathematics,* **4** (1991), 99–106.

[12] N. C. Wormald, Models of random regular graphs, in: *Surveys in Combinatorics, 1999,* London Mathematical Society Lecture Note Series 267 (J. D. Lamb and D. A. Preece, eds), Cambridge University Press (Cambridge, 1999), pp. 239–298.

Noga Alon
Tel Aviv University
Tel Aviv 69978
Israel
and
IAS, Princeton, NJ
08540, USA
e-mail: nogaa@tau.ac.il

Nicholas Wormald
Department of Combinatorics and Optimization
University of Waterloo
Waterloo ON
Canada
e-mail: nwormald@uwaterloo.ca

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 23–42.

ITERATED TRIANGLE PARTITIONS

STEVE BUTLER* and RON GRAHAM

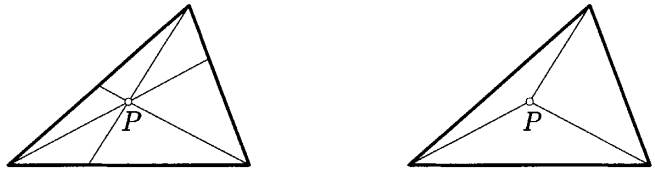
For a given triangle there are many points associated with the triangle that lie in its interior; examples include the incenter (which can be found by the intersection of the angle bisectors) and the centroid (which can be found by the intersection of the medians). Using this point one can naturally subdivide the triangle into either three or six "daughter" triangles. We can then repeat the same process on each of the daughters and so on and so on. A natural question is after some large number of steps what does a typical nth generation daughter look like (up to similarity)? In this paper we look at this problem for both the incenter and the centroid and show that they have very distinct behavior as n gets large. We will also consider the Gergonne point and the Lemoine point.

1. INTRODUCTION

In this paper we will be considering triangles and are concerned only with the shape, i.e., up to similarity. So we can denote a triangle T by the triple $T(A, B, C)$ where A, B and C are the interior angles. Given a point P in the interior of T we can use the point to subdivide T into smaller "daughter" triangles. This is done by taking each vertex and drawing the line passing through P and connecting it to the opposite side (sometimes referred to as Cevians), subdividing the triangle into six daughters (Figure 1a); or by taking each vertex and drawing the line to P subdividing into three daughters (Figure 1b).

We are interested in the case when P is a well-defined point lying in the interior of T for any non-degenerate T. Examples of such points include the

*This work was partially done with support of an NSF Mathematical Sciences Postdoctoral Fellowship.



(a) Six daughters (b) Three daughters

Fig. 1. Given an interior point P how to subdivide the triangle

incenter (which is the intersection of the angle bisectors of T), the *centroid* (which is the intersection of the medians of T), along with many others. A rather complete listing of well known distinguished points associated with triangles is maintained at the Encyclopedia of Triangle Centers [9] (however not all of these points will always lie in the interior of T, the excenter being one such example).

Once we have settled on a way of choosing P, we can iterate the process of subdivision, say for n times, producing 6^n (or 3^n) nth generation daughters from our original triangle T. We are interested in what can be said about an nth generation daughter of T as n goes to infinity. For example, what is the distribution of the shape of the daughters? What is the distribution of the smallest angle (or second smallest angle) of T? It will turn out that the answers to these questions depend in a crucial (and currently not well understood) way on exactly how P is chosen. In the following sections, we will specify various choices for P and address these questions and show that well known points P can produce dramatically different results as n goes to infinity.

To help us answer these questions we first will need a convenient way of describing the triangles. Each triangle $T(A, B, C)$ corresponds to a point $(A, B, C) \in \mathbb{E}^3$ which will lie in the intersection of the plane $x + y + z = \pi$ with the positive orthant (see Figure 2). We will denote this intersection, which is actually an equilateral triangle, by P (this representation has been used before in the analysis of pedal triangles; see [1, 10, 11]).

2. Subdividing by Bisectors

One of the easiest points P to work with is the incenter which is formed by using the angle bisectors. What is nice about this point is that the angles in

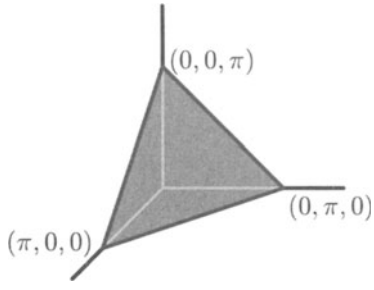


Fig. 2. Representing triangles as points in $\mathbb{E}^3$

the daughter triangles are linear combinations of the angles in the original triangle. So the angles of the daughter triangle can be formed via matrix multiplication. This transforms the problem into a Markov process using matrices which has been extensively studied (see the survey [6] for more information). We will consider both possibilities of subdivision shown in Figure 1.

2.1. Subdividing into three daughters

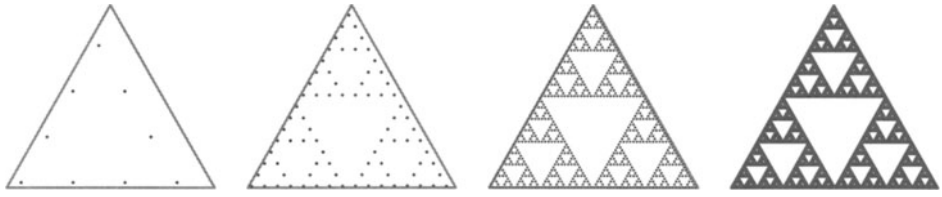
If we associate the triangle $T(A,B,C)$ with the vector $\mathbf{t} = (A,B,C)^*$ in $\mathbb{E}^3$, then the three daughters of T can be found by $M_i\mathbf{t}$ where $i \in \{1,2,3\}$ and

$$M_1 = \begin{pmatrix} 1/2 & 0 & 0 \\ 1/2 & 1 & 1/2 \\ 0 & 0 & 1/2 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1 & 1/2 & 1/2 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1/2 & 0 \\ 1/2 & 1/2 & 1 \end{pmatrix}.$$

The nth generation daughters of T can then be found by $M_{i_1}M_{i_2}\cdots M_{i_n}\mathbf{t}$ where each $i_j \in \{1,2,3\}$. Starting with the triangle $T(\pi/9, 2\pi/9, 2\pi/3)$ we have plotted the nth generation daughters for $n = 2,4,6,8$ in Figure 3.

The figures are very suggestive about what is happening. To see why we are getting the Sierpinski triangle it helps to see how the image of M_i maps P to P. This is shown in Figure 4.

So we have that there will be exactly one daughter in each of the subtriangles indicated in Figure 5a. Similarly, if we look at the location of the second generation daughters there will be one daughter in each of the nine triangles indicated in Figure 5b, and if we look at the location of the third generation daughters there will be one daughter in each of the twenty seven triangles indicated in Figure 5c. Since this process mirrors that



(a) $n = 2$ (b) $n = 4$ (c) $n = 6$ (d) $n = 8$

Fig. 3. The nth generation daughters for angle bisectors with three daughters at each generation

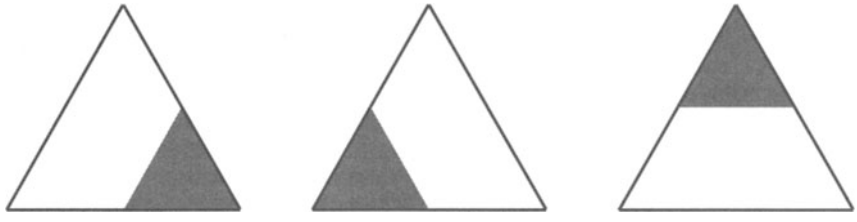
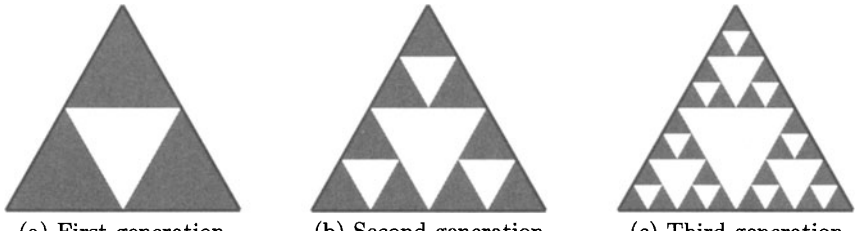


Fig. 4. The image of P under the three maps M_i

used to form the Sierpinski triangle, as n goes to infinity the distribution of daughters are the points in the Sierpinski triangle.



(a) First generation (b) Second generation (c) Third generation

Fig. 5. Location of the descendants in the first three generations

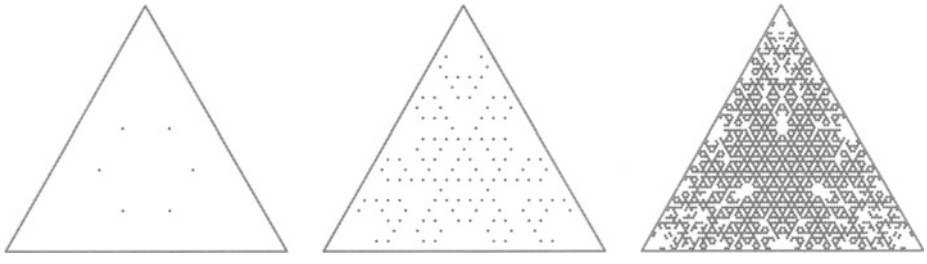
2.2. Dividing into six triangles

As before if we associate a triangle $T(A, B, C)$ with the vector $\mathbf{t} = (A, B, C)^*$ in $\mathbb{E}^3$, then the six daughters of T can be found by $M_i\mathbf{t}$ where $i \in \{1, 2, 3, 4, 5, 6\}$ and

$$M_1 = \begin{pmatrix} 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 0 \\ 1/2 & 0 & 1 \end{pmatrix}, \quad M_3 = \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \end{pmatrix},$$

$$M_4 = \begin{pmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 \end{pmatrix}, \quad M_5 = \begin{pmatrix} 1/2 & 0 & 1/2 \\ 1/2 & 1 & 0 \\ 0 & 0 & 1/2 \end{pmatrix}, \quad M_6 = \begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

And as before we can form the nth generation daughters by looking at all 6^n possible products of the form $M_{i_1} M_{i_2} \cdots M_{i_n} \mathbf{t}$ with each $i_j \in \{1, 2, 3, 4, 5, 6\}$. Starting with the equilateral triangle $T_1 = T(\pi/3, \pi/3, \pi/3)$ we have plotted the nth generation daughters for $n = 1, 3, 5$ in Figure 6.



(a) $n = 1$ (b) $n = 3$ (c) $n = 5$

Fig. 6. The nth generation daughters for angle bisectors with six daughters at each generation

Looking at Figure 6c we see that the daughters seem to fill in P rather uniformly. A patient count though will reveal that there are far fewer than 6^5 triangles in Figure 6c. It is not hard to see that an nth generation daughter of the equilateral triangle must have the form $T(u_1\pi/2^n, u_2\pi/2^n, u_3\pi/2^n)$ for positive integers u_1, u_2, u_3 with $u_1 + u_2 + u_3 = 2^n$. In particular there are at most $\binom{2^n}{2}$ possible different daughter triangles among the 6^n which will be generated. So on average each triangle is being hit approximately $2(3/2)^n$ times. In particular, any triangle which is not being hit must be missed for a good reason. It is not too hard to see that one condition that is needed is for $\gcd(u_1, u_2, u_3) = 1$, which rules out a positive fraction of the daughters but does not still explain all of the missing daughters.

If we start with a triangle which does not have rational multiples of π then we do not have as many triangles stacking up on top of each other. As an example, the fifth generation daughters of $T_2 = T\big(\pi/3, \sqrt{2}\pi/3, \big(2 - \sqrt{2}\big)\pi/3\big)$ are shown in Figure 7 (this should be compared to Figure 6c).

One problem that we run into is that the triangles can be fairly densely packed in these pictures. So instead of plotting the individual triangles it is better to look at a histogram. We will divide P into a large number of small

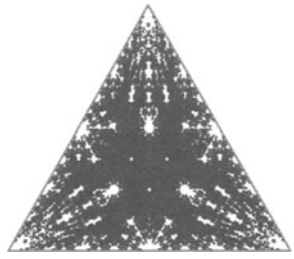


Fig. 7. The fifth generation daughters for angle bisectors starting with an "irrational" triangle

regions and then shade each region according to the number of triangles that fall into that region, the darker a region is the more triangles fall into that region. We have plotted histograms for T_1 and T_2 in Figure 8 (these should be compared with Figures 6c and 7).

The situation is obviously much richer than the previous case when we looked at three daughters. The reason for this is previously the image of each one of the M_i was disjoint, but this is no longer the case. If we plot the images of the six maps we get Figure 9, in particular there is a lot of overlap between pairs of maps.

However, we still have one nice feature of these maps. Namely, they are contracting. To see this we note that we can put P into $\mathbb{E}^2$ by putting $T(A,B,C)$ at $\big((A+2B)/\sqrt{3},A\big)$. We then can put the first generation daughter $T\big(A/2,(A+B)/2,(B+2C)/2\big)$ at $\big((3A+2B)/\big(2\sqrt{3}\big),A/2\big)$. If we now compare the distance between triangles $T(A,B,C)$ and $T(A',B',C')$ and their daughters, a calculation shows

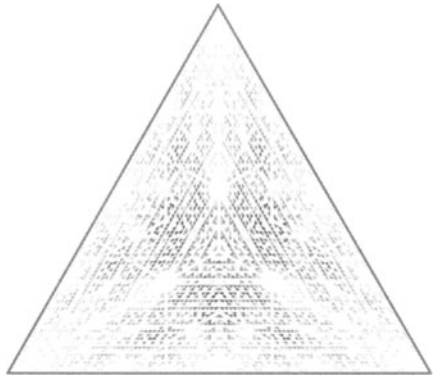
$$\frac{3}{4}\left(\left(\frac{A+2B}{\sqrt{3}}-\frac{A'+2B'}{\sqrt{3}}\right)^2+(A-A')^2\right)$$

$$-\left(\left(\frac{3A+2B}{2\sqrt{3}}-\frac{3A'+2B'}{2\sqrt{3}}\right)^2+\left(\frac{A}{2}-\frac{A'}{2}\right)^2\right)=\frac{2}{3}(B-B')^2\geq 0.$$

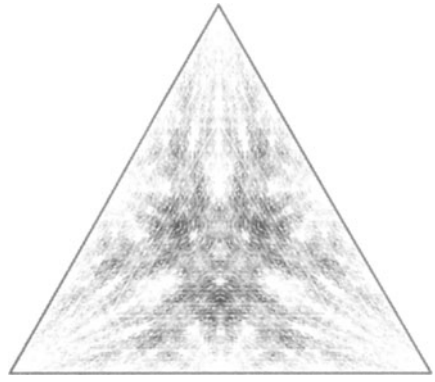
It follows that for the euclidean distance d we have

$$d\big(T\big(A/2,(A+B)/2,(B+2C)/2\big),T\big(A'/2,(A'+B')/2,(B'+2C')/2\big)\big)$$

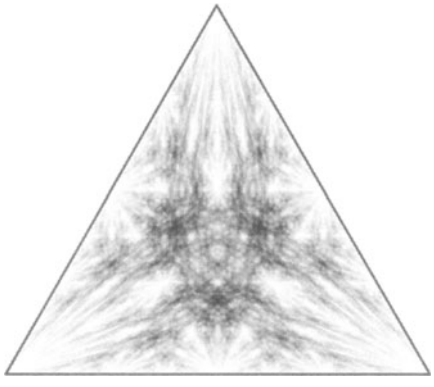
$$\leq\frac{\sqrt{3}}{2}d\big(T(A,B,C),T(A',B',C')\big).$$



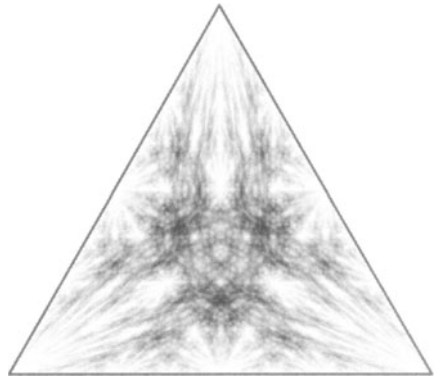
(a) Triangle T_1; $n = 5$



(b) Triangle T_2; $n = 5$



(c) Triangle T_1; $n = 12$



(d) Triangle T_2; $n = 12$

Fig. 8. Histograms for the distribution of triangles in P under bisector division

Fig. 9. The image of P under the six maps M_i

By symmetry, the same statement holds for all the daughter triangle maps.

Since these maps are contracting with Lipschitz constant $\sqrt{3}/2$ then it follows (see [6]) that there is a fixed stationary distribution on P that the process converges to. Further it converges exponentially. Hence Figures 8c and 8d are nearly identical, and these are approximations for the histogram of the limiting distribution.

2.2.1. The smallest angle. We now consider the problem of analyzing the distributions of the smallest angles. Just as we have a quick rate of convergence to the stationary distribution in P; we experimentally see a quick convergence to the same distribution of minimal angles. In Figure 10 we split the interval between 0 and $\pi/3$ into 1000 intervals and took 50,000,000 random walks of length 50 in this Markov process and recorded the minimum angle.

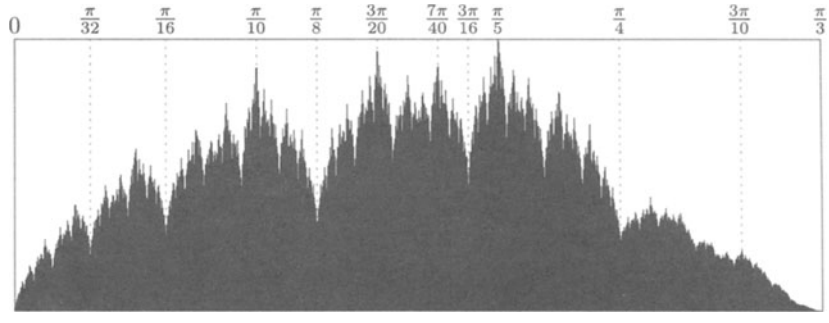


Fig. 10. The distribution of the smallest angle using bisectors

A few things certainly stand out. Perhaps the simplest thing to observe is that the minimum angles are not concentrated near 0, so that the nth generation daughters are not becoming "flat". There also is a certain amount of self-similarity (but given the distributions we saw in Figure 8cd this should not be surprising). We also have marked a few of the values between 0 and $\pi/3$ in Figure 10. In particular we see that a few of the "valleys" are located at $\pi/16, \pi/8, 3\pi/16, \pi/4$ and a few of the peaks are located at $\pi/10, 3\pi/20, 7\pi/40, \pi/5$.

The tallest peak is located at $\pi/5$. If we look closely at the histograms in Figure 8cd we see that the darkest region is located at $T(\pi/5, \pi/5, 2\pi/5)$. This particular triangle has the following unique property; $T(\pi/5, 2\pi/5, 2\pi/5)$ is the only triangle where *two* of its daughters are similar to itself (this is shown as the shaded triangles in Figure 11). In terms of the M_i this means that some permutation of the vector $\mathbf{t} = (\pi/5, 2\pi/5, 2\pi/5)^*$ is an eigenvector associated with eigenvalue 1 for two of the M_i. The only other non-degenerate triangle that also acts as an eigenvector is $T(2\pi/9, \pi/3, 4\pi/9)$ which has *one* of its daughters similar to itself.

Figure 10 was computed using random walks and so the distribution is at best approximate. It is not too difficult to find upper and lower bounds for the distribution (though in practice the amount of computation needed

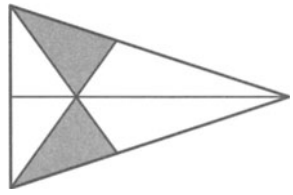


Fig. 11. The decomposition of $T(\pi/5, 2\pi/5, 2\pi/5)$

to compute these bounds to an accuracy to be able to produce Figure 10 is prohibitive).

The basic idea is to note for a fixed k that $M_{i_1} \cdots M_{i_k}$ maps P into a smaller triangular region of P, let us denote this by Q. So in particular we have that $M_{i_1} \cdots M_{i_k} M_{i_{k+1}} \cdots M_{i_n} \mathbf{t}$ must lie in Q as well for any choice of initial $\mathbf{t}$ and any choice of the $i_{k+1}, \ldots, i_n$. In essence if we know what the last k steps of this Markov process are then we have a lot of information about where in P we end up (the higher the value of k the more precisely we can approximate the point). With this in mind the procedure is to fix a value of k and then look at the image of P under each of the 6^k possible combinations of $M_{i_1} \cdots M_{i_k}$. Now divide the interval between 0 and $\pi/3$ in some fashion. For each interval a lower bound for the distribution is found by counting the number of images Q which *must* have smallest angle in that interval divided by 6^k; an upper bound is found by counting the number of images of Q which *can* have some smallest angle in the interval divided by 6^k. An example of the resulting bounds on the distribution is shown for $k = 15$ and the interval split into widths of $\pi/256$ in Figure 12.

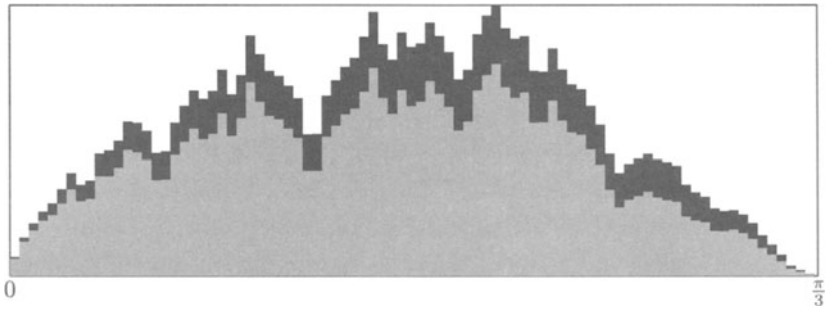


Fig. 12. Upper (black) and lower (red) bounds for the true distribution of smallest angle

The precise location of the peaks and valleys is in general not well understood. For comparison we also include the distribution of the middle

and largest angles in Figures 13 and 14 computed in the same way as Figure 10.

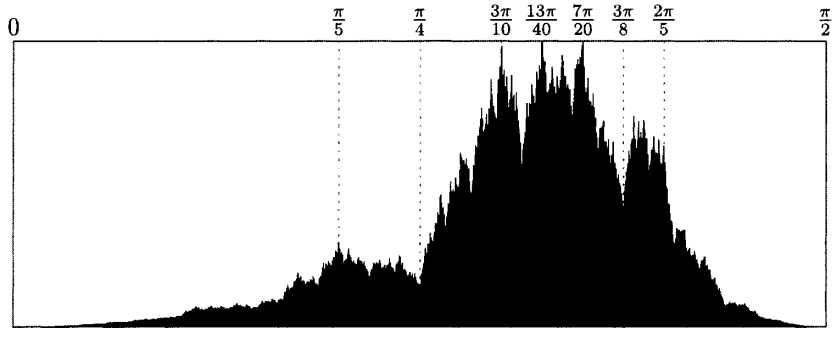


Fig. 13. The distribution of the middle angle using bisectors

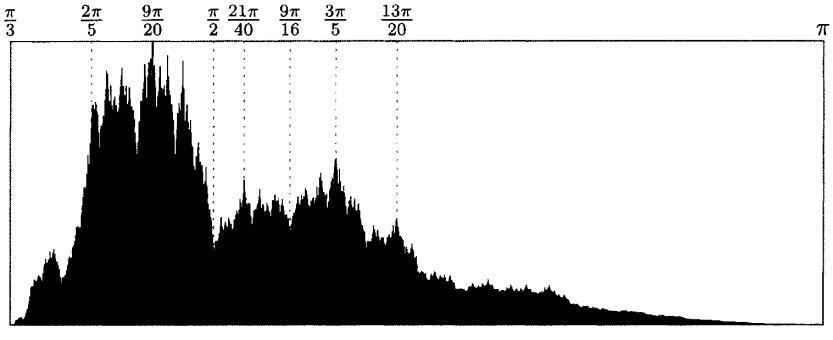


Fig. 14. The distribution of the largest angle using bisectors

3. Subdividing by Medians

We now consider the subdivision using the centroid, or median point of the triangle. We will restrict our attention to the case of six daughters. This case has been well studied because of its close relationship to barycentric subdivision. We will see that this situation is quite different from the case of the incenter. One major difference is that the median map is not contracting on P as we saw previously. In Figure 15 we indicate how one of the median daughters maps P to itself (the other five are similar and differ by rotation and reflections).

To get some intuition about what is happening in this case let us consider the histogram for $T(\pi/3, \pi/3, \pi/3)$ for $n = 4, 6, 8, 10$ shown in Figure 16.

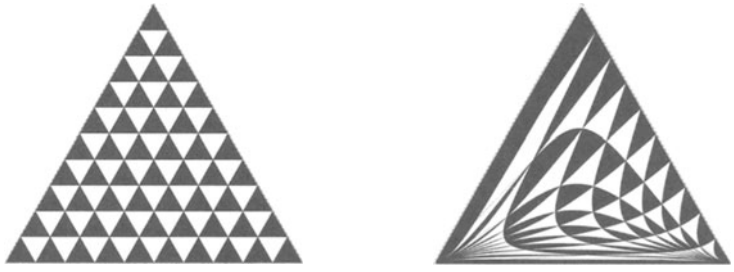
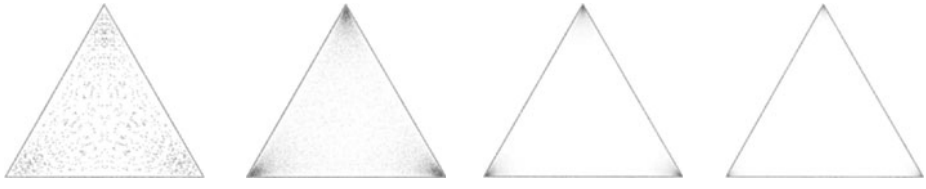


Fig. 15. A subdivision of P and its image under the median map



(a) $n = 4$ (b) $n = 6$ (c) $n = 8$ (d) $n = 10$

Fig. 16. The nth generation daughters for medians with six daughters at each generation

Looking at these pictures what seems to be happening is that the triangles in P tend to go towards the vertices as n increases. Or in other words, the triangles are becoming "flat", or nearly colinear, as n gets large. This behavior was noted by David Blackwell and was passed on to us by Persi Diaconis [5], independently Bárány et al. [2, 15], inspired by a question of Stakhovskii about the distribution of shapes of triangles, showed that most triangles are flat.

3.1. Analytic approach

Blackwell started by putting a triangle in a standard position by putting the longest side along the x-axis with vertices at $(0,0)$ and $(1,0)$, the third vertex is then put with $y > 0$ and $x \geq 1/2$. Given a triangle T in standard position we now define the *pseudo-fatness* of T, denoted $PF(T)$, by $PF(T) = \sqrt{y}(3 - x)$. It should be noted that for a triangle in standard position that $y/2$ is the area of the triangle, so the term $\sqrt{y}$ relates to the area. The term $(3 - x)$ acts as an error correction term.

If $T_1, \ldots, T_6$ are the six daughter triangles of T then Blackwell noted that if the average of the pseudo-fatness of the daughters was small compared to

the pseudo-fatness of T for *each* triangle T, i.e.,

$$\frac{1}{6}\sum_{T_i} PF(T_i) \leq c\,PF(T)$$

for all T and some $c < 1$, then most of the triangles will have pseudo-fatness going to zero. In particular, most of the triangles would be flat.

Computation using Maple confirms this assertion for the value $c = 0.99$ as we show in Figure 17. From this, for example, it would follow that the number of nth generation daughters with smallest angle greater than 0.99^n radians is at most 5.9^n. However, at present no one has confirmed analytically that the above inequality is valid with this value of c. Also it would be interesting to find the best pseudo-fatness function.

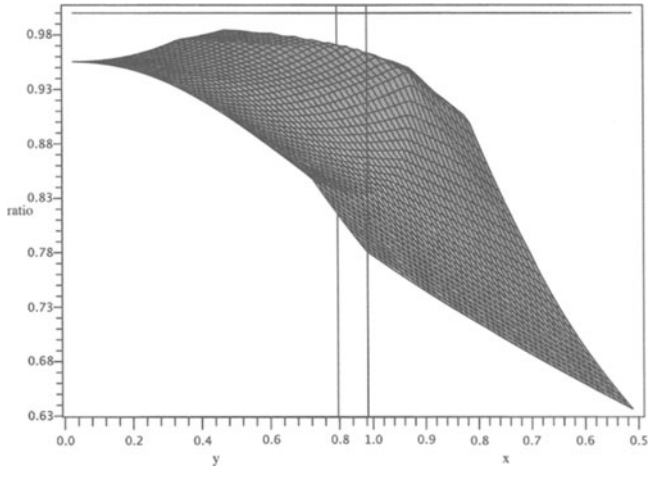


Fig. 17. Average of the pseudo-fatness of daughter triangles, seen from the side

There are weaknesses in this approach. Namely, while we can show that the smallest angle is small, this does not automatically imply that the largest angle is large ($\approx \pi$). Or put another way, this approach shows that the triangles drift to the edges of P but not necessarily to the vertices of P. Robert Hough [8] using a different technique was able to show that the largest angle does in fact approach π and moreover gave asymptotic bounds for the proportion of triangles with angles near π, this shows that the triangles do accumulate at the vertices of P.

3.2. Hyperbolic approach

The method of Bárány et al. was to move the problem from P to the hyperbolic half plane. The following approach is similar to theirs and was discovered independently by Curt McMullen [14]. The first step is to associate with each triangle T (up to) six points z in the hyperbolic upper half plane $\mathbb{H}$ as shown in Figure 18. Namely, *some* edge of T is located with vertices at $z = 0$ and $z = 1$ and the third vertex is located at the complex coordinate z with positive imaginary part. McMullen then observes that reflecting z across the three circles $\Re(z) = \frac{1}{2}$, $|z| = 1$, and $|z - 1| = 1$ induces a natural action of S_3 on $\mathbb{H}$ in which all six orientations of T occur.

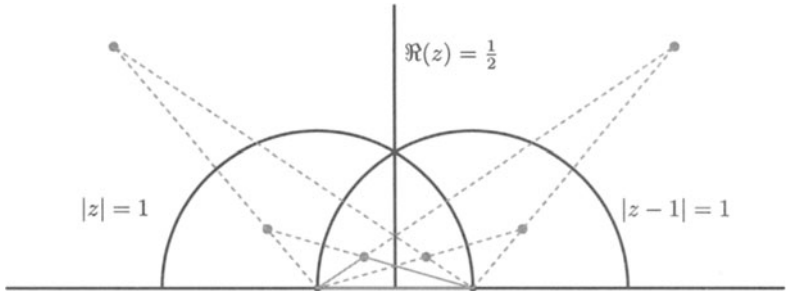


Fig. 18. Triangles as points in the hyperbolic plane

Now, the centroid point of a triangle in this position is represented by the point $(z + 1)/3$, and so one of the corresponding daughters becomes $2(z + 1)/3$ when normalized. The group of automorphisms of $\mathbb{H}$ generated by the map $B(z) = 2(z + 1)/3$ and S_3 is dense in $Aut(\mathbb{H})$ (see [2]). From this, using results of Furstenberg [7], it follows that almost all random walks formed from products of $B(z)$ and elements of S_3 tend to infinity (in the hyperbolic plane) as the length of the product increases. This then implies that almost all of the nth generation daughters have smallest angle tending to 0 as n increases. Actually more can be said in that these points are dense in the hyperbolic space and so every triangle is arbitrarily "close" so some descendant of the starting triangle.

4. Subdividing by the Gergonne Point

We again restrict ourselves to the case of six daughters. The Gergonne point of a triangle is the point of concurrence of the three line segments joining each vertex of the triangle to the point of tangency of the inscribed circle to the side opposite the vertex. We show an example of this in Figure 19. (As a side note the example shown is the only triangle (up to similarity) with the equilateral triangle as one of the daughters; its side lengths are proportional to 19, 40 and 49, see [4].)

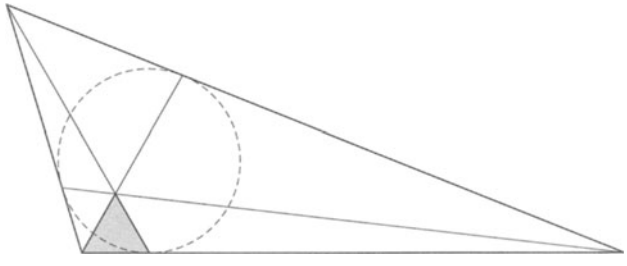


Fig. 19. A triangle and its Gergonne point G

In Figure 20 we look at how one of the Gergonne daughters maps P to itself. It is interesting to note that unlike the medians or angle bisectors this map is not 1-to-1. That is, there is a small region which gets mapped 2-to-1.

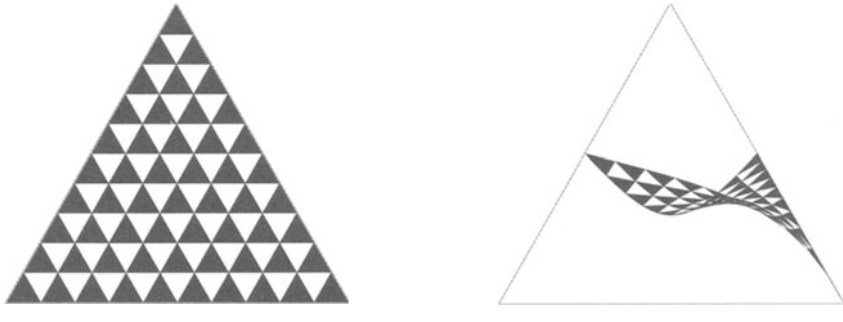
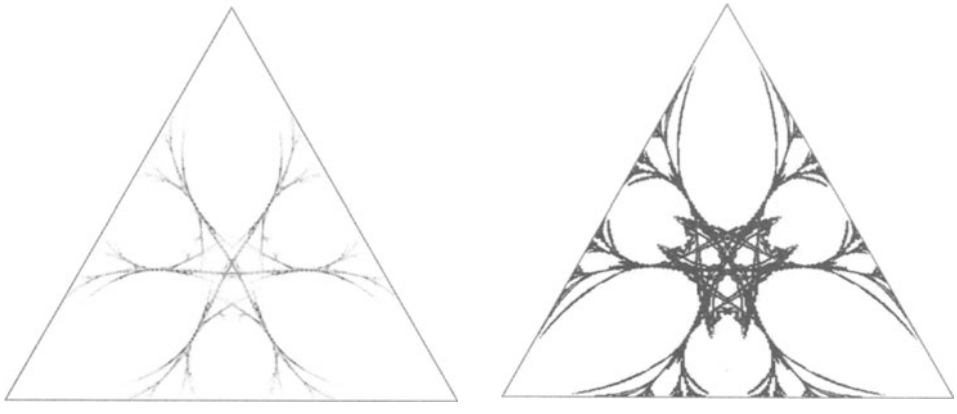


Fig. 20. A subdivision of P and its image under the Gergonne map

Looking at Figure 20 it also looks like the map might be contracting (this would then imply a limiting distribution). However this is not the case. For instance if we consider the following two triangles in the corner $T_1(170°, 5°, 5°)$ and $T_2(160°, 10°, 10°)$ then two of their corresponding

daughters are respectively $DT_1(170.0377\ldots^\circ, 9.9245\ldots^\circ, 0.0377\ldots^\circ)$ and $DT_2(160.2935\ldots^\circ, 19.4129\ldots^\circ, 0.2935\ldots^\circ)$. So $\|DT_1 - DT_2\|/\|T_1 - T_2\| = 1.1106\ldots$ showing that there is a slight expansion in the corner. This does not rule out a limiting distribution (and the authors do believe that a limiting distribution exists) it only shows that more sophisticated techniques will be needed.

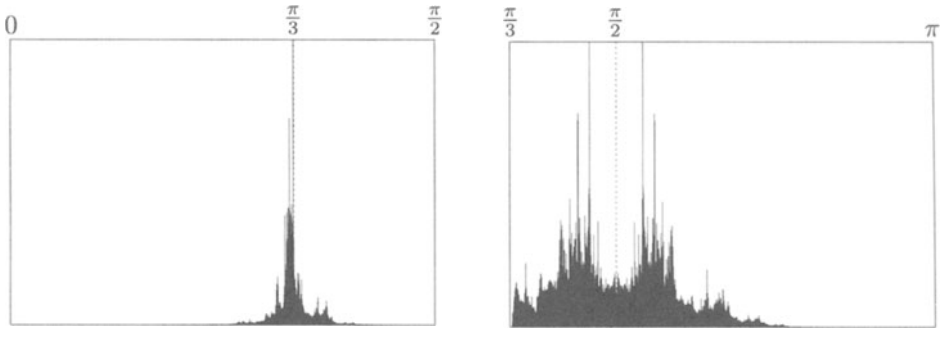
In Figure 21a we have drawn the histogram for the tenth generation daughters (higher generation daughters seem to have a similar structure). There appears to be a fair amount of white, this is caused not by there being very few triangles in the region, but *no* triangles in the region. In Figure 21b we have darkened the histogram by making any region with a daughter black and any region without a daughter white. As the figure indicates there seems to be large regions with missing daughters. It would be interesting to know what triangles are possible for the nth generation (this is not even known precisely for the first generation).



(a) Histogram (b) Darkened

Fig. 21. The tenth generation daughters for the Gergonne point with six daughters at each generation

In Figure 22 we indicate the distributions for the middle and largest angles using the Gergonne point divison. Looking at the distributions it seems that the middle angle is concentrated around $\pi/3$ while the largest angle has some approximate symmetry around $\pi/2$. For the largest angle what appears to be happening is that for "most" triangles when using the Gergonne point the largest angles of the daughters are located at the vertex along the edge of the parent triangle. So by pairing up adjacent daughters along the edge their two largest angles should be symmetric around $\pi/2$.

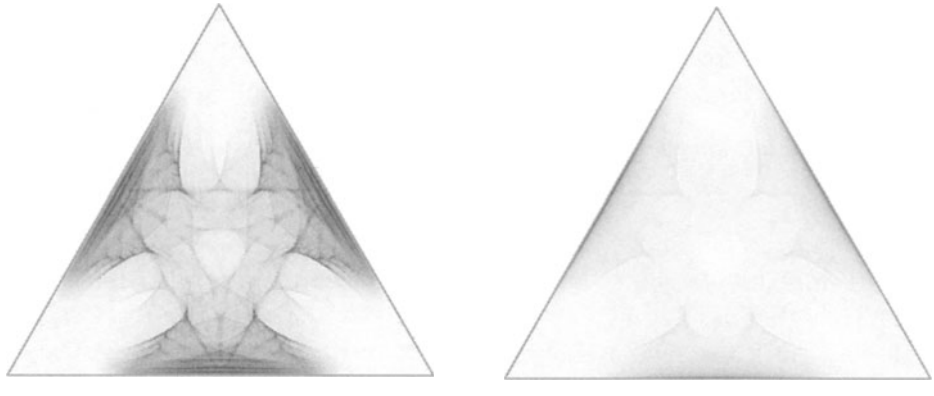


(a) Middle angle (b) Largest angle

Fig. 22. The distribution of angles using the Gergonne point

5. Subdividing by the Lemoine Point

The Lemoine point of a triangle is the intersection of the symmedians which are the medians reflected across the angle bisectors, so it is also known as the symmedian point. This point seems to be similar to the medians in that the smallest angles go to 0 but experimentally the triangles seem to be going to the *sides* of P rather than the *corners* of P. Further, they seem to be drifting *slowly*. In Figure 23a we have plotted the histogram for all the 10th generation daughters while Figure 23b was formed by random walks of length 30 using the Lemoine point to subdivide.



(a) $n = 10$ (b) $n = 30$

Fig. 23. The nth generation daughters for the Lemoine point with six daughters at each generation

In Figure 24 we look at how one of the Lemoine daughters maps P to itself. Clearly the map is not contracting. It is also difficult to see what is happening on the right hand side, this is because there is a region which gets mapped 3-to-1 onto. It would be interesting to know more about this mapping, for instance where is the folding occurring (experimentally it seems to be folding near the preimage point of $T(\pi/3, \pi/3, \pi/3)$).

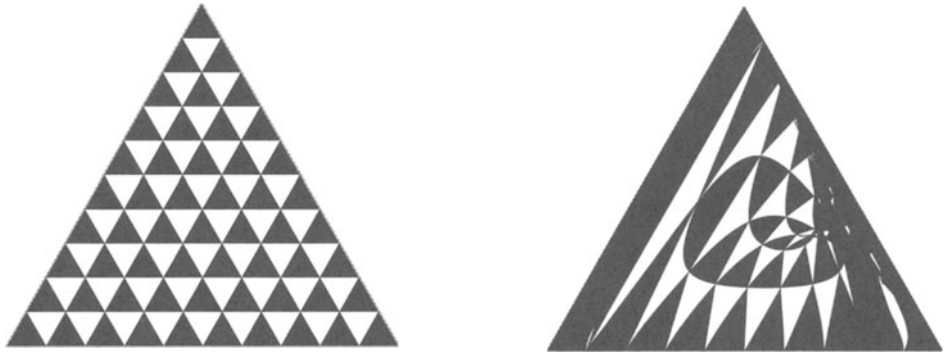


Fig. 24. A subdivision of P and its image under the Lemoine map

6. Concluding Remarks

There are still a large number of questions yet to be answered. For instance, what can be said about the limiting distribution when using bisectors? Is there a limiting distribution when using the Gergonne point, if so what is it? What is the behavior as n tends to infinity when using the Lemoine point? If the triangles are becoming flat what can be said about the rate of convergence?

When looking at the medians we were interested in how the two smallest angles compared. When looking at the square of the smallest angle divided by the second smallest angle (giving a ratio between 0 and 1) we got the histogram in Figure 25 divided into 5000 slots; 13611504/17702781 went into the first slot and so we removed that extreme case. The remaining data showed spikes in unusual places, which seem to be connected to the Farey fractions. Is there any explanation for this behavior?

In addition, there is of course a vast catalogue of additional points that can be considered (see [9]). One can also consider variations such as alternating between bisectors and medians at each stage, or using as

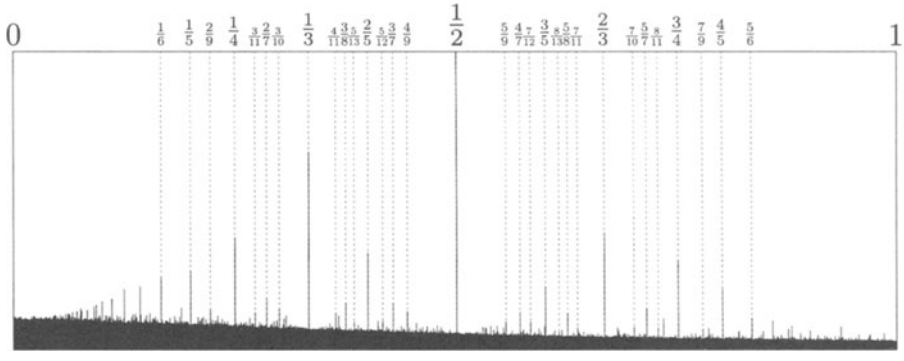


Fig. 25. The square of the ratio of the smallest and middle angles using 4091277 random walks with medians of depth 300

a central point the midpoint between the incenter and centroid. A much more general question would be to ask what characterizes points where in the limit most triangles are not flat?

One could also consider the problem of where we pick the point uniformly at random in each daughter. In this direction Mannion [12, 13] has shown that if we choose all three vertices at random in the triangle then almost surely the limiting process is collinear.

Similar questions can be asked for higher dimensional analogues. For example Schwartz [18, 19] looked at the distribution of "shapes" when n-simplices are recursively subdivided using Barycentric subdivision. In that case he showed that the behavior is similar to the case of division of triangles using the centroid. It would be interesting to see if there were other well defined points in simplices which generate different behavior (i.e., similar to that of the incenter or Gergonne points).

Beyond the intrinsic curiosity of these iterated triangle partitions there is also a vast literature on how to subdivide triangles so that the minimal angle is always bounded away from 0 (see [3, 16, 17]). This is desirable since certain methods can fail when the subdivision creates a large number of triangles with minimal angles going to 0 as n gets large. So for instance using medians for subdividing is not desirable since this creates many flat triangles. On the other hand, if we use angle bisectors we can guarantee that all but $f(\varepsilon)$ proportion of the triangles have minimal angles at least ε so that we can limit the proportion of "bad" triangles.

Acknowledgements. We thank Persi Diaconis and Curt McMullen for many useful discussions and insights and the referee for directing us to related literature.

REFERENCES

[1] J. C. Alexander, The symbolic dynamics of pedal triangles, *Math. Mag.*, **66** (1993), 147–158.

[2] Imre Bárány, Alan F. Beardon and Thomas Keith Carne, *Barycentric subdivision of triangles and semigroups of Möbius maps*, Mathematika **43** (1996), 165–171.

[3] Marshall Bern and David Eppstein, "Mesh generation and optimal triangulation", in: *Computing in Euclidean Geometry*, Ding-Zhu Du and Frank Hwang, eds. (1992), 23–90.

[4] Steve Butler, The lost daughters of Gergonne, *Forum Geometricorum*, **9** (2009), 19–26.

[5] Persi Diaconis, personal communication.

[6] Persi Diaconis and David Freedman, Iterated random functions, *SIAM Review*, **41** (1999), 45–76.

[7] Harry Furstenberg, Noncommuting random products, *Trans. Amer. Math. Soc.*, **108** (1963), 377–428.

[8] Robert Hough, personal communication.

[9] Clark Kimberling, Encyclopedia of Triangle Centers, `http://faculty.evansville.edu/ck6/encyclopedia/ETC.html`

[10] John Kingston and John Synge, The sequence of pedal triangles, *Amer. Math. Monthly*, **95** (1988), 609–620.

[11] Peter Lax, The ergodic character of sequences of pedal triangles, *Amer. Math. Monthly*, **97** (1990), 377–381.

[12] David Mannion, Convergence to collinearity of a sequence of random triangle shapes, *Adv. Appl. Prob.*, **22** (1990), 831–844.

[13] David Mannion, The invariant distribution of a sequence of random collinear triangle shapes, *Adv. Appl. Prob.*, **22** (1990), 845–865.

[14] Curt McMullen, personal communication.

[15] Andrei A. Ordin, A generalized barycentric subdivision of a triangle, and semigroups of Möbius transformations, *Russian Math. Surveys*, **55** (2000), 591–592.

[16] María-Cecilia Rivar and Gabriel Iribarren, The 4-triangles longest-side partition of triangles and linear refinement algorithms, *Math. of Computation*, **65** (1996), 1485–1502.

[17] Ivo Rosenberg and Frank Stenger, A lower bound on the angles of triangles constructed by bisecting the longest side, *Math. of Computation*, **29** (1975), 390–395.

[18] Richard Evan Schwartz, The density of shapes in three-dimensional barycentric subdivision, *Discrete Comput. Geom.*, **30** (2003), 373–377.

[19] Richard Evan Schwartz, Affine subdivision, steerable semigroups, and sphere coverings, *Pure and Applied Mathematics Quarterly*, **3** (2007), 897–926.

Steve Butler

Department of Mathematics
UCLA, Los Angeles
CA 90095

e-mail: butler@math.ucla.edu

Ron Graham

Department of Computer Science and Engineering
University of California
San Diego, La Jolla
CA 92093

e-mail: graham@ucsd.edu

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 43–62.

PageRank and Random Walks on Graphs

FAN CHUNG and WENBO ZHAO

Dedicated to Lovász on the ocassion of his sixtieth birthday

We examine the relationship between PageRank and several invariants occurring in the study of random walks and electrical networks. We consider a generalized version of hitting time and effective resistance with an additional parameter which controls the 'speed' of diffusion. We will establish their connection with PageRank. Through these connections, a combinatorial interpretation of PageRank is given in terms of rooted spanning forests by using a generalized version of the matrix-tree theorem. Using PageRank, we will illustrate that the generalized hitting time leads to finding sparse cuts and efficient approximation algorithms for PageRank can be used for approximating hitting time and effective resistance.

1. Introduction

The notion of PageRank, first introduced by Brin and Page [2], forms the basis for their Web search algorithms. Although the original version of PageRank was used for the Webgraph (with all the webpages as vertices and hyperlinks as edges), PageRank is well defined for any given graph and is quite effective for capturing various relations among vertices of graphs. In this paper, we will investigate several implications of PageRank for a given graph.

To start with, we give the graph-theoretical definition of PageRank. Roughly speaking, PageRank is a way to organize random walk of various lengths. Instead of having to determine the number of steps a random walks is taking, PageRank uses a positive real value α, where $\alpha \in [0, 1)$ to control the "diffusion" of a combination of random walks.

The original definition for PageRank was to assign a value to each vertex (Webpage), denoting the "importance" of a vertex under two assumptions: For some fixed probability α, a surfer at a Webpage jumps to a random Webpage with probability α and goes to a linked Webpage with probability $1-\alpha$. The importance of a Webpage v is the expected sum of the importance of all the Webpages u that preceed v.

In this paper, we will use a more general version of PageRank, called *personalized PageRank*, introduced by Jeh and Widom [7] (also see Haveliwala [6]). The personalized PageRank $\mathrm{pr}_\alpha(s)$ depends on two parameters, the *jumping constant* α and a *seed* s. A seed can be viewed as a vertex or a probability distribution on vertices. The original definition of PageRank is the special case where the seed is the uniform distribution.

To define the PageRank for a connected graph G, we consider random walks on G with transition probability matrix P and the lazy random walk on G, denoted by $Z = (I + P)/2$. In this paper, all vectors are regarded as row vectors unless stated otherwise.

The personalized PageRank vector $\mathrm{pr}_\alpha(s)$ with a jumping constant α and a seed vector s is defined to be the unique solution of the linear system

$$\mathrm{pr}_\alpha(s) = \alpha s + (1-\alpha)\,\mathrm{pr}_\alpha(s)Z. \tag{1}$$

An alternate but equivalent definition for pr_α is an expression of a geometric sum of random walks:

$$\mathrm{pr}_\alpha(s) = \alpha s \sum_{k=0}^{\infty} (1-\alpha)^k Z^k.$$

In addition to the practical applications of Websearch algorithms, PageRank has numerous connections to various graph invariants. For example, PageRank can be used to find cuts with a certain isoperimetric guarantee, similar to the Cheeger inequalities. It was shown in [1] that for an arbitrarily chosen vertex u, if we arrange vertices in a row using PageRank $\mathrm{pr}_\alpha(u)$, one of the cuts which consists of vertices in initial segments has Cheeger ratio optimal up to a quadratic factor with high probability. The performance guarantee is quite similar to that given by the spectral partitioning algorithm using eigenvectors. However, the advantages of using PageRank are multifold. We can choose an appropriate α to specify the approximate size of the part that we wish to cut and thus PageRank leads to so-called *local* algorithms. Furthermore, there are effective algorithms for computing

approximate PageRank with finite support with size depending only on the error bound and the desired size of the smaller separated part [1].

In this paper, we explore the relationship of PageRank and various graph invariants occuring in random walks and electrical networks. In the spirit of PageRank, we consider a generalized version of the hitting time and effective resistance with an additional parameter α. In a way, these generalized invariants provide a quantative ranking of edges, indicating how important an edge is while allowing the choice of α. In Section 2, we will define the Laplacian and the discrete Green's function as well as their connection to PageRank. In Section 3, we consider electrical networks and the generalized versions of hitting time and the effective resistance. In Section 4, we will give several matrix-forest theorems, which generalize the classical Matrix-Tree Theorem [11]. In Section 5, we will derive a combinatorial interpretation of PageRank in terms of spanning forests in the graph. In Section 6, we consider some useful properties of the generalized hitting time in connection of identifying sparse cuts. In Section 7, we use PageRank to estimate the effective resistance.

2. Laplacian, the Green's Function and PageRank

We consider a connected weighted undirected graph $G = (V, E, w)$. Suppose G has vertex set V, edge set E, edge weight $w_{u,v} \geq 0$ and $|V| = n$, $|E| = m$. A typical random walk is determined by the transition probabilities $P(u, v) = w_{uv}/d_u$ where the degree d_u of u is the sum $\sum_{v:\{u,v\}\in E} w_{u,v}$. The volume of a subset $S \subseteq V$, denoted by $\mathrm{vol}\,(S)$, is the sum of degrees of vertices in S. In particular, the volume of G, denoted by $\mathrm{vol}\,(G)$, is equal to $\mathrm{vol}\,(V)$. For the special case of $w_{u,v} = 1$ for all $\{u, v\} \in E$, we have $\mathrm{vol}\,(G) = 2m$.

Let A denote the weighted adjacency matrix with entries $A(u, v) = w_{u,v}$ and D denote the diagonal degree matrix. Then the transition probability matrix P is equal to $D^{-1}A$ and for any initial distribution f, the distribution of the random walk after k steps is fP^k. For undefined terminology, the reader is referred to the excellent survey of Lovász [13]

The combinatorial Laplacian of G is defined by $L = D - A$. If we orient the edges of G in an arbitrary but fixed way, we can write its Laplacian as

$$L = B^T W B, \tag{2}$$

where B is the signed edge-vertex incidence matrix, given by

$$B(e,v) = \begin{cases} 1 & \text{if } v \text{ is } e\text{'s head} \\ -1 & \text{if } v \text{ is } e\text{'s tail} \\ 0 & \text{otherwise} \end{cases}$$

and W is the digonal matrix with $W(e,e) = w_e$. The normalized Laplacian of G is defined to be $\mathcal{L} = D^{-1/2}LD^{-1/2}$ and we can write

$$\mathcal{L} = S^T W S$$

where $S = BD^{-1/2}$.

Since $\mathcal{L}$ is symmetric and we can express $\mathcal{L}$ by

$$\mathcal{L} = \sum_{i=0}^{n-1} \lambda_i \phi_i^T \phi_i = \sum_{i=1}^{n-1} \lambda_i \phi_i^T \phi_i,$$

where $\lambda_0 = 0$ and $0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n-1} \leq 2$ are the nonzero eigenvalues of $\mathcal{L}$ and $\phi_0, \ldots, \phi_{n-1}$ form a corresponding orthonormal basis of eigenvectors. The fact of $\lambda_1 > 0$ follows from the connectivity of G. The eigenvalue λ_1 is intimately related to the rate of convergence of random walks. The reader is referred to [4] for numerous properties concerning eigenvalues of the normalized Laplacian. Although the combinatorial Laplacian is useful for various flow problems in the study of electrical networks, the spectrum of combinatorial Laplacian is not effective (except for almost regular graphs) for applications requiring isoperimetric properties.

Denote the β-normalized Laplacian $\mathcal{L}_\beta$ by $\beta I + \mathcal{L},$. Then we may write $\mathcal{L}_\beta = S'^T W_\beta S'$ where we define S' and W_β as follows:

$$S' = \begin{bmatrix} I \\ S \end{bmatrix}_{(n+m)\times n} \quad B' = \begin{bmatrix} D^{1/2} \\ B \end{bmatrix}_{(n+m)\times n} \quad \text{and } W_\beta = \begin{bmatrix} \beta I & 0 \\ 0 & W \end{bmatrix}_{(n+m)\times(n+m)}.$$

For simplicity, we index the columns of S' and the columns of W_β by $V \cup E$ where the first n columns are indexed by V and the last m columns are indexed by E. The rows of W_β are indexed in the same way. It is easy to verify that

$$\mathcal{L}_\beta = D^{-1/2} B'^T W_\beta B' D^{-1/2} = S'^T W_\beta S'.$$

Green's functions were first introduced in a celebrated essay by George Green [8] in 1828. Since then, the concept of Green's functions has been

used in a wide range of areas, especially in the study of partial differential equations and quantum field theory. The discrete analog of Green's functions, which are associated with the normalized Laplacian of graphs, were first introduced in a 2002 paper [5] in connection with the study of Dirichlet eigenvalues with boundary conditions. The Green's function $\mathcal{G}$ denotes the symmetric matrix satisfying

$$f\mathcal{L}\mathcal{G} = f\mathcal{G}\mathcal{L} = f.$$

for all vectors f which are orthogonal to the eigenvector

$$\phi_0 = \mathbf{1}D^{1/2}/\sqrt{\mathrm{vol}\,(G)}$$

where $\mathbf{1}$ denotes the all 1's vector. The Green's function $\mathcal{G}$ has the following form:

$$\mathcal{G} = \sum_{i=1}^{n-1} \frac{1}{\lambda_i} \phi_i^T \phi_i. \tag{3}$$

The following modified Green's function $\mathcal{G}_\beta$ was also used in [5]. For $\beta \in \mathbb{R}^+$, let Green's function $\mathcal{G}_\beta$ denote the symmetric matrix satisfying

$$\mathcal{L}_\beta \mathcal{G}_\beta = I. \tag{4}$$

Clearly, we have

$$\mathcal{G}_\beta = \sum_{i=0}^{n-1} \frac{1}{\lambda_i + \beta} \phi_i^T \phi_i. \tag{5}$$

By comparing with the recurrence of the PageRank in (1), we remark that the discrete Green's function is basically a symmetric form of the PageRank. Namely, we can write

$$\frac{\mathrm{pr}_\alpha(s)}{\beta} = sD^{-1/2}\mathcal{G}_\beta D^{1/2} \tag{6}$$

where

$$\beta = \frac{2\alpha}{1-\alpha}. \tag{7}$$

3. PageRank, the Hitting Time and the Effective Resistance in Electrical Networks

A connected weighted undirected graph $G = (V, E, w)$ can be viewed as an electrical network, where the edge weight w_e represents the *conductance* of e. The *effective resistance* $R(u, v)$ between two vertices u and v is defined as the voltage potential difference induced between them when a unit current is injected at u and extracted at v. The effective resistance can be characterized by the combinatorial Laplacian of the graph (see [11]). Since we wish to establish the connection of the effective resistance with PageRank, we will consider the normalized Laplacian instead.

Suppose we are given the injected current function $i_V : V \to \mathbb{R}$. The induced current i_E on the edges satisfies the property that the sum of all induced current on edges entering v is equal to $i_V(v)$, as asserted by *Kirchoff's current law:*

$$i_V = i_E B.$$

For any function $f : V \to \mathbb{R}$, we can regard f as a *voltage potential* function in the following sense: The induced flow for the edge from u to v is the product of $f(u) - f(v)$ and the conductance of the edge, according to *Ohm's law,* and can be expressed by

$$i_E = f_V B^T W.$$

We can write

$$i_V D^{-1/2} = (f_V B^T W) B D^{-1/2} = f_V D^{1/2} \mathcal{L}.$$

Suppose we only consider the voltage potential function f satisfying $\sum_v f(v) = 0$. By using the definition of Green's function, we have

$$i_V D^{-1/2} \mathcal{G} D^{-1/2} = f_V.$$

Suppose we inject a unit current to vertex u and extract a unit current from v, i.e., $i_V = \chi_v - \chi_u$ where χ_u is the characteristic function with $\chi_u(x) = 1$ if $x = u$ and 0 otherwise. Thus, the effective resistance between vertices u and v can be written as

$$\begin{aligned} R(u, v) &= f_V(\chi_v - \chi_u)^T \\ &= i_V D^{-1/2} \mathcal{G} D^{-1/2} (\chi_v - \chi_u)^T \end{aligned} \tag{8}$$

$$= (\chi_v - \chi_u)D^{-1/2}\mathcal{G}D^{-1/2}(\chi_v - \chi_u)^T.$$

The effective resistance is closely associated with the hitting time and commute time for a random walk on G. The hitting time $H(u,v)$ is the expected number of steps of a random walk starting from u until it first visit v. The commute time $C(u,v)$ of u and v is the expected number of steps in a random walk starting at u, before vertex v is visited, and the vertex u is reached again. The commute time can be expressed as the sum

$$C(u,v) = H(u,v) + H(v,u).$$

It has been shown [12] (also see [3]) that the commute time $C(u,v)$ satisfies

$$C(u,v) = \frac{R(u,v)}{\mathrm{vol}\,(G)}.$$

For two distinct vertices u and v, the hitting time $H(u,v)$ satisfies the following equation:

$$H(u,v) = 1 + \frac{1}{d_u} \sum_{w\,:\,\{w,u\}\in E} H(w,v) \tag{9}$$

and $H(u,u) = 0$. Here we will express the hitting time in terms of the Green's function.

Lemma 1. *For all $u, v \in V$,*

$$\frac{H(u,v)}{\mathrm{vol}\,(G)} = (\chi_v - \chi_u)D^{-1/2}\mathcal{G}D^{-1/2}\chi_v^T.$$

Proof. Clearly, $H(v,v) = 0$ for all $v \in V$. We use the equation in (9). Consider

$$\frac{1}{\mathrm{vol}\,(G)} + \frac{1}{d_u} \sum_{w\,:\,(w,u)\in E} (\chi_v - \chi_u)D^{-1/2}\mathcal{G}D^{-1/2}\chi_v^T$$

$$= \frac{1}{\mathrm{vol}\,(G)} + \chi_v D^{-1/2}\mathcal{G}D^{-1/2}\chi_v^T - \chi_u D^{-1}AD^{-1/2}\mathcal{G}D^{-1/2}\chi_v^T$$

$$= \frac{1}{\mathrm{vol}\,(G)} + \chi_v D^{-1/2}\mathcal{G}D^{-1/2}\chi_v^T - \chi_u D^{-1/2}(I - \mathcal{L})\mathcal{G}D^{-1/2}\chi_v^T$$

$$= (\chi_v - \chi_u)D^{-1/2}\mathcal{G}D^{-1/2}\chi_v^T$$

$$+ \chi_u D^{-1/2}\left(\frac{1}{\mathrm{vol}\,(G)}D^{1/2}JD^{1/2} + \mathcal{L}\mathcal{G}\right)D^{-1/2}\chi_v^T$$

$$= (\chi_v - \chi_u)D^{-1/2}\mathcal{G}D^{-1/2}\chi_v^T$$

$$+ \chi_u D^{-1/2}\left(\frac{1}{\mathrm{vol}\,(G)}D^{1/2}JD^{1/2} + I - u_0^T u_0\right)D^{-1/2}\chi_v^T$$

$$= (\chi_v - \chi_u)D^{-1/2}\mathcal{G}D^{-1/2}\chi_v^T + \chi_u D^{-1}\chi_v^T$$

$$= (\chi_v - \chi_u)D^{-1/2}\mathcal{G}D^{-1/2}\chi_v^T \qquad \blacksquare$$

We define

$$h(u,v) = \frac{H(u,v)}{\mathrm{vol}\,(G)}.$$

Now, we define a generalized hitting time with an additional parameter $\alpha > 0$.

$$h_\alpha(u,v) = \frac{\left[\mathrm{pr}_\alpha(\chi_v)\right](v)}{d_v} - \frac{\left[\mathrm{pr}_\alpha(\chi_v)\right](u)}{d_u}. \tag{10}$$

We define the generalized effective resistance $R_\alpha(u,v)$ as follows:

$$R_\alpha(u,v) = h_\alpha(u,v) + h_\alpha(v,u).$$

It is easy to check that $R_\alpha(u,v)$ satisfies

$$R_\alpha(u,v) = \beta(\chi_v - \chi_u)D^{-1/2}\mathcal{G}_\beta D^{-1/2}(\chi_v - \chi_u)^T$$

where β satisfies (7). By using the connection with PageRank in (6), we can write

$$R_\alpha(u,v) = \frac{\left[\mathrm{pr}_\alpha(\chi_u)\right](u)}{d_u} - \frac{\left[\mathrm{pr}_\alpha(\chi_u)\right](v)}{d_v} + \frac{\left[\mathrm{pr}_\alpha(\chi_v)\right](v)}{d_v} - \frac{\left[\mathrm{pr}_\alpha(\chi_v)\right](u)}{d_u}. \tag{11}$$

4. Several Matrix-Forest Theorems

It is well-known that in a simple graph G for an edge joining u and v, the effective resistance $R(u,v)$ is proportional to the number of spanning trees containing the edge $\{u,v\}$. Later on we will show that the generalized effective resistance $R_\alpha(u,v)$ is proportional to a combination of spanning forests of certain types. A *forest* is a graph containing no cycle. A k-rooted forest is a forest with k connected components where each of the connected components contains a special vertex that we call a root. A *k-rooted spanning forest* is a k-rooted forest containing all vertices as vertices. It has exactly $n-k$ edges. The special case for $k=1$ is a rooted spanning tree.

The weight of a forest F is defined to be the product of all weights of edges in F. We define the weight $w(F)$ of an unrooted forest as follows:

$$w(F) = \prod_{e \in F} w(e).$$

For example, if F is a rooted tree in a simple graph, then $w(F) = 1$. For a rooted tree F, we denote by F_* the set of all vertices in F which are not roots. The weight $w^*(F)$ of a rooted forest is defined by

$$w^*(F) = \frac{\prod_{e \in F} w(e)}{\prod_{v \in F_*} d_v}.$$

First, we prove several useful facts along the same lines as the Matrix-Tree Theorem [11]

Theorem 1. *In a graph $G = (V, E, w)$, for $\beta > 0$, the determinant of $\mathcal{L}_\beta$ is related to rooted spanning forests as follows:*

$$\det \mathcal{L}_\beta = \sum_{k=1}^{n} \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)$$

where $\mathcal{F}_k$ denotes the family of all k-rooted spanning forests in G.

Proof. For a matrix M and subsets X, Y of indices of rows and columns, we denote a submatrix $M_{X,Y}$ of M by restricting rows and columns of M to X and Y. We consider the determinant of $\mathcal{L}_\beta$ as follows:

$$\begin{aligned} \det \mathcal{L}_\beta &= \det S'^T W_\beta S' \\ &= \sum_{\substack{F \subseteq V \cup E \\ |F|=n}} (\det S'_{F,V})(\det W_{F,F})(\det S'_{F,V}) \end{aligned}$$

by the Cauchy–Binet Theorem. We note that $\det S'_{F,V} = 0$ if F contains a cycle. In the other direction, if F does not contain a cycle, F is a spanning forest and the vertices in $F \cap V$ are the k roots of the spanning forest. Therefore, we have

$$\det \mathcal{L}_\beta = \sum_{k=1}^{n} \sum_{\substack{F \\ k\text{-rooted spanning forest} \\ \text{with roots in } F \cap V}} \left(\det S'_{F,V}\right)^2 \det\left(W_\beta\right)_{F,F}.$$

Note that for a k-rooted spanning forest F, we have

$$\det S'_{F,V} = \pm \frac{\beta^k}{\sqrt{\prod_v d_v}} \prod_{u \in F \cap V} \sqrt{d_u}.$$

Thus we have

$$\det \mathcal{L}_v = \frac{1}{\prod_v d_v} \sum_{k=1}^{n} \beta^k \sum_{\substack{F \\ k\text{-rooted spanning forests}}} \prod_{u \in F \cap V} d_u \prod_{e \in F} w_e$$

$$= \sum_{k=1}^{n} \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)$$

as claimed. ■

Theorem 2. *In a graph $G = (V, E, w)$, for $\beta > 0$ and a vertex v in G, the determinant of the principle minor $\mathcal{L}_\beta^{(v)}$, obtained by excluding the row and column associated with v satisfies:*

$$\det \mathcal{L}_\beta^{(v)} = \sum_{k=1}^{n} \beta^{k-1} \sum_{F \in \mathcal{F}_{k,v}} w^*(F)$$

where $\mathcal{F}_{k,v}$ denotes the family of k-rooted trees having v as one of the roots.

Proof. Let V' denote $V \setminus \{v\}$. We can write the determinant of $\mathcal{L}_\beta^{(v)}$ as follows:

$$\det \mathcal{L}_\beta^{(v)} = \det\left(\left(S'_{V' \cup E, V'}\right)^T W_\beta S'_{V' \cup E, V'}\right)$$

$$= \sum_{\substack{F \subseteq V' \cup E \\ |F \cap V'| = k-1 \\ |F| = n-1}} \left(\det S'^*_{F,V'}\right)\left(\det S'_{F,V'}\right) \det\left(W_\beta\right)_{F,F}$$

$$= \sum_{k=1}^{n} \sum_{\substack{F \\ k\text{-rooted spanning forest} \\ \text{with roots in } F \cap V' \text{ and } v}} \left(\det S'_{F,V'}\right)^2 \det\left(W_\beta\right)_{F,F}$$

since if the forest F satisfies $\det S'_{F,V} \neq 0$, then $F \subseteq V' \cup E$ and F forms a k-rooted spanning forest with v as a root. Thus,

$$\det S'_{F,V'} = \pm \frac{\beta^{k-1}}{\sqrt{\prod_{u \neq v} d_u}} \prod_{u \in F \cap V} d_u,$$

and we have

$$\det \mathcal{L}_\beta^{(v)} = \frac{1}{\prod_{u \neq v} d_u} \sum_{k=1}^{n} \beta^{k-1} \sum_{\substack{F \\ k\text{-rooted spanning forests} \\ v \text{ is a root}}} \prod_{u \in F \cap V} d_u \prod_{e \in F} w_e$$

$$= \sum_{k=1}^{n} \beta^{k-1} \sum_{F \in \mathcal{F}_{k,v}} w^*(F)$$

as desired. ■

Theorem 3. *In a graph $G = (V, E, w)$, for $\beta > 0$ and two distinct vertices u, v in G, the determinant of the minor $\mathcal{L}_\beta^{(u,v)}$, obtained by excluding the row associated with u and column associated with v satisfies:*

$$\det \mathcal{L}_\beta^{(u,v)} = \sqrt{\frac{d_v}{d_u}} \sum_{k=1}^{n} \beta^{k-1} \sum_{F \in \mathcal{F}_{k,u,v}} w^*(F)$$

where $\mathcal{F}_{k,u,v}$ denotes the family of k-rooted spanning forests in which u and v are in the same connected component and u is a root.

Proof. We denote $V_1 = V \setminus \{u\}$ and $V_2 = V \setminus \{v\}$. We consider the determinant of $\mathcal{L}_\beta^{(u,v)}$ as follows:

$$\det \mathcal{L}_\beta^{(u,v)} = \sigma(u,v) \det\left(\left(S'_{V_1 \cup E,V}\right)^T W_\beta S'_{V_2 \cup E,V}\right)$$

$$= \sigma(u,v) \sum_{k=1}^{n} \sum_{\substack{F \subseteq V \cup E \\ |F \cap V_1| = k-1 \\ |F| = n-1}} \left(\det S'_{F,V_1}\right)\left(\det S'_{F,V_2}\right) \det\left(W_\beta\right)_{F,F}$$

$$= \sigma(u,v) \sum_{k=1}^{n} \sum_{\substack{F \subseteq V \cup E \\ |F \cap V''| = k-1 \\ |F| = n-1}} \left(\det S'_{F,V_1}\right) \left(\det S'_{F,V_2}\right) \det\left(W_\beta\right)_{F,F}$$

since if u or v is in F, one of the determinants in the product of the preceding equality becomes zero. Here $\sigma(u,v)$ denotes $(-1)^{i+j}$ if u is the index of the ith row and v is the index of the jth column of $\mathcal{L}_\beta$. If $\det S'_{F,V_1}$ is non-zero, the edges in F form a spanning tree with roots $\{u\} \cup (F \cap V'')$ where $V'' = V \backslash \{u,v\}$. If $\det S'_{F,V_2}$ is non-zero, then the edges in F form a spanning tree with roots $\{v\} \cup (F \cap V'')$. Therefore if the product is nonzero, vertices u and v are in the same connected component in the spanning forest formed by the edges in F. Suppose $F \subseteq V'' \cup E$ forms a rooted spanning forest with k connected components, one of which contains both u and v. Thus we have

$$\det \mathcal{L}_\beta^{(u,v)} = \frac{\sigma(u,v)}{\sqrt{d_u d_v} \prod_{x \neq u,v} d_x} \sum_{k=1}^{n} \sum_{\substack{F \subseteq V'' \cup E \\ |F \cap V''| = k-1 \\ |F| = n-1}}$$

$$\times \left(\det B'_{F,V_1}\right) \left(\det B'_{F,V_2}\right) \det\left(W_\beta\right)_{F,F}$$

We will prove the following Claim later.

Claim.

$$\det B'_{F,V_2} = \sigma(u,v) \det B'_{F,V_1}.$$

From the Claim, we have

$$\det \mathcal{L}_\beta^{(u,v)} = \frac{1}{\sqrt{d_u d_v} \prod_{x \neq u,v} d_x} \sum_{k=1}^{n} \sum_{\substack{F \\ k\text{-rooted spanning forest} \\ u \text{ is a root} \\ u,v \text{ are in the same c.c.}}}$$

$$\times \left(\det B'_{F,V_1}\right)^2 \det\left(W_\beta\right)_{F,F}$$

$$= \frac{1}{\sqrt{d_u d_v} \prod_{x \neq u,v} d_x} \sum_{k=1}^{n} \sum_{F \in \mathcal{F}_{k,u,v}} \left(\det B'_{F,V_1}\right)^2 \det\left(W_\beta\right)_{F,F}$$

$$= \frac{1}{\sqrt{d_u d_v} \prod_{x \neq u,v} d_x} \sum_{k=1}^{n} \sum_{F \in \mathcal{F}_{k,u,v}} \beta^{k-1} \prod_{e \in F \cap E} w_e \prod_{x \in F \cap V} d_x$$

$$= \frac{1}{\sqrt{d_u d_v}} \sum_{k=1}^{n} \beta^{k-1} \sum_{F \in \mathcal{F}_{k,u,v}} \frac{\prod_{e \in F} w_e}{\prod_{x \in F_* \setminus \{v\}} d_x}$$

$$= \sqrt{\frac{d_v}{d_u}} \sum_{k=1}^{n} \beta^{k-1} \sum_{F \in \mathcal{F}_{k,u,v}} w^*(F).$$

It remains to prove the Claim.

Proof of the Claim. We prove the claim in three steps:

Case 1: First we see that the Claim holds if F forms a path with vertices $u = u_1, u_2, \ldots, u_{k+1} = v$ and edges $e_i = \{u_i, u_{i+1}\}$ in F where u_i is the index for the ith row of $\mathcal{L}$, and e_j is the index of the jth column in B.

Case 2: Since u and v are in the same connected component of F, let $u = v_1, v_2, \ldots, v_t = v$ denote the path joining u and v in F. If u_i is the index for the ith row of $\mathcal{L}$, and e_j is the index of the jth column in B for $1 \le i \le t+1$ and $1 \le j \le t$, then the claim holds in the same way as in Case 1.

Case 3: Now we consider the general case. We follow the notation in Case 2 concerning the paths joining u and v in F. We denote by U the permutation on V that moves v_i to the ith place for $i = 1, \ldots, k+1$ and denote by U' the permutation on E that moves $e'_j = \{v_j, v_{j+1}\}$ to the jth place for $j = 1, \ldots, k$. We consider

$$\sigma(u,v) \det\left(\left(B'_{F,V_1}\right)^T B'_{F,V_2}\right)$$

$$= \sigma\big(U(u), U(v)\big) \det\left(U {B'_{F,V_1}}^T U'^T U' B'_{F,V_2} V\right)$$

$$= \sigma\big(U(u), U(v)\big) \det\left(U'^T B'_{F,V_1} U\right) \det\left(U' B'_{F,V_2} V\right)$$

We have reduced Case 3 to Case 2. Thus the above expression is equal to 1 and the Claim is proved.

The proof of Theorem 3 is completed. ■

5. PageRank and Other Invariants in Terms of Rooted Spanning Forests

Using the matrix-forest theorems in the preceding section, we can establish the connection between the discrete Green's function and rooted spanning forests. We consider the discrete Green's function $\mathcal{G}_\beta$ for $\beta > 0$. For a fixed vertex v, we consider

$$\mathcal{L}_\beta f = \chi_v. \tag{12}$$

Clearly, $f = \mathcal{G}_\beta \chi_v$ is a solution to the above equation. We can view (12) as a linear system with variables $f(u)$ for u in V. By Cramer's rule, we have

$$f(v) = \mathcal{G}_\beta(v,v) = \frac{\det \mathcal{L}_\beta^{(v)}}{\det \mathcal{L}_\beta} \tag{13}$$

$$f(u) = \mathcal{G}_\beta(u,v) = \frac{\det \mathcal{L}_\beta^{(u,v)}}{\det \mathcal{L}_\beta}. \tag{14}$$

From equation (6) and Theorems 1, 2 and 3, we can express PageRank in terms of the combinatorial sums involving rooted spanning forests as follows:

Theorem 4. *For a vertex v in G, the PageRank pr_α satisfies*

$$\left[\mathrm{pr}_\alpha(\chi_v)\right](v) = \frac{\sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_{k,v}} w^*(F)}{\sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}$$

where $\beta = 2\alpha/(1-\alpha)$, $\mathcal{F}_k$ denotes the family of all k-rooted spanning forests in G and $\mathcal{F}_{k,v}$ denotes the family of k-rooted trees having v as one of the roots.

Theorem 5. *For two distinct vertices u and v in G, the PageRank pr_α satisfies*

$$\left[\mathrm{pr}_\alpha(\chi_u)\right](v) = \frac{d_v \sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_{k,v}} w^*(F)}{d_u \sum_{k=1}^n \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}$$

where $\beta = 2\alpha/(1-\alpha)$ and $\mathcal{F}_{k,u,v}$ denotes the family of k-rooted spanning forests in which u and v are in the same connected component and u is a root.

As it turns out, the generalized hitting time can also be expressed in a similar form:

Theorem 6. *For an edge $e = \{u, v\}$ in G, the hitting time $h_\alpha(u, v)$ satisfies*

$$h_\alpha(u,v) = \frac{\beta \sum_{k=1}^{n-1} \beta^k \sum_{F \in \mathcal{F}_{k,e,v}} w^*(F)}{w_e \sum_{k=1}^{n} \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}$$

where $\beta = 2\alpha/(1-\alpha)$ and $\mathcal{F}_{k,e,v}$ denotes the family of k-rooted spanning forests containing e as an edge which is in the path from v to its root.

Proof. From (10), we have

$$\begin{aligned} h_\alpha(u,v) &= \frac{\left[\mathrm{pr}_\alpha(\chi_v)\right](v)}{d_v} - \frac{\left[\mathrm{pr}_\alpha(\chi_v)\right](u)}{d_u} \\ &= \frac{\sum_{k=1}^{n} \beta^k \left(\sum_{F \in \mathcal{F}_{k,v}} w^*(F) - \sum_{F \in \mathcal{F}_{k,v,u}} w^*(F) \right)}{d_v \sum_{k=1}^{n} \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)} \end{aligned}$$

Note that the set of $\mathcal{F}_{k,v} \setminus \mathcal{F}_{k,v,u}$ consists of k-rooted trees in which u and v are in different connected components and v is a root. For every rooted forest F in $\mathcal{F}_{k,v} \setminus \mathcal{F}_{k,v,u}$ we consider the rooted tree F' formed by adding the edge $e = \{u, v\}$ where v is no longer a root. The path from v to its root in F' contains e. Clearly, F' has $k-1$ connected components and $k-1$ roots. Let $\mathcal{F}_{k-1,e,v}$ denote the family of $k-1$ forests with the additional property that the path from v to its root contains e. Thus F' is in $\mathcal{F}_{k-1,e,v}$. We note

$$w^*(F') = \frac{w_e}{d_v} w^*(F)$$

since v is no longer a root. This is a bijection from $\mathcal{F}_{k,v} \setminus \mathcal{F}_{k,v,u}$ to $\mathcal{F}_{k-1,e,v}$. We can then write

$$\begin{aligned} h_\alpha(u,v) &= \frac{\sum_{k=1}^{n} \beta^k \sum_{F' \in \mathcal{F}_{k-1,e,v}} w^*(F')/w_e}{\sum_{k=1}^{n} \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)} \\ &= \frac{\beta \sum_{k=1}^{n-1} \beta^k \sum_{F \in \mathcal{F}_{k,e,v}} w^*(F)}{w_e \sum_{k=1}^{n} \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}. \end{aligned}$$

The proof is complete. ■

By using the above theorems, we can then derive a relatively clean formulation for the generalized effective resistance.

Theorem 7. *For an edge $e = \{u, v\}$ in G and $\alpha \in [0, 1)$, the effective resistance $R_\alpha(u, v)$ satisfies*

$$R_\alpha(u,v) = \frac{\beta \sum_{k=1}^{n-1} \beta^k \sum_{F \in \mathcal{F}_{k,e}} w^*(F)}{w_e \sum_{k=1}^{n} \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}$$

where $\beta = 2\alpha/(1-\alpha)$ and $\mathcal{F}_{k,e}$ denotes the family of k-rooted forests containing e as an edge with k roots.

Proof. The proof follows from the fact that $\mathcal{F}_{k,e,u}$ and $\mathcal{F}_{k,e,v}$ are disjoint and

$$\mathcal{F}_{k,e} = \mathcal{F}_{k,e,u} \cup \mathcal{F}_{k,e,v}.$$

Therefore we have

$$\begin{aligned} R_\alpha(u,v) &= h_\alpha(u,v) + h_\alpha(v,u) \\ &= \frac{\beta \sum_{k=1}^{n-1} \beta^k \sum_{F \in \mathcal{F}_{k,e}} w^*(F)}{w_e \sum_{k=1}^{n} \beta^k \sum_{F \in \mathcal{F}_k} w^*(F)}. \end{aligned}$$

■

6. Using the Generalized Hitting Time to Find Sparse Cuts

There are several useful properties of the hitting time h_α which we will describe.

Lemma 2. *For a subset $S \subset V$ with $S \neq \emptyset$, we have*

$$\sum_{\substack{\{u,v\} \in E \\ u \in S, v \notin S}} h_\alpha(u,v) \leq \frac{2\alpha}{1-\alpha}.$$

Proof. Since we can view $\mathrm{pr}_\alpha(\chi_v)$ as a probability distribution, we have

$$\begin{aligned} \sum_{\substack{\{u,v\} \in E \\ u \in S, v \notin S}} h_\alpha(u,v) &= \sum_{\substack{\{u,v\} \in E \\ u \in S, v \notin S}} \left(\frac{[\mathrm{pr}_\alpha(\chi_v)](v)}{d_v} - \frac{[\mathrm{pr}_\alpha(\chi_v)](u)}{d_u} \right) \\ &= [\mathrm{pr}_\alpha(\chi_v)](S) - [\mathrm{pr}_\alpha(\chi_v)P](S). \end{aligned}$$

Since the PageRank vector $\mathrm{pr}_\alpha(\chi_v)$ is the unique solution of the following equation,

$$\mathrm{pr}_\alpha(\chi_v) = \alpha\chi_v + (1-\alpha)\,\mathrm{pr}_\alpha(\chi_v)\frac{I+P}{2},$$

by rearranging the terms above, we have

$$\mathrm{pr}_\alpha(\chi_v) - \mathrm{pr}_\alpha(\chi_v)P = \frac{2\alpha}{1-\alpha}\chi_v - \frac{2\alpha}{1-\alpha}\,\mathrm{pr}_\alpha(\chi_v).$$

Thus,

$$\begin{aligned}\left[\mathrm{pr}_\alpha(\chi_v)\right](S) - \left[\mathrm{pr}_\alpha(\chi_v)P\right](S) &= \frac{2\alpha}{1-\alpha}\chi_v(S) - \frac{2\alpha}{1-\alpha}\left[\mathrm{pr}_\alpha(\chi_v)\right](S)\\ &\le \frac{2\alpha}{1-\alpha}. \qquad \blacksquare\end{aligned}$$

For two distinct vertices, we consider a partition of V into two parts, one of which contains u and the other contains v. In particular, we are interested in the Cheeger ratio of such partitions defined as follows:

$$\phi_{uv} = \min_{\substack{S\subset V\\ u\in S, v\in\bar{S}}} \frac{e(S,\bar{S})}{\min\left\{\mathrm{vol}\,(S), \mathrm{vol}\,(\bar{S})\right\}}.$$

Theorem 8. *For two distinct vertices u and v and a constant $\phi \in (0,1)$, suppose two sets $X = \{w : h_\alpha(w,v) \le 0\}$ and $Y = \{w : h_\alpha(w,v) < h_\alpha(u,v)\}$ satisfy the following conditions:*

(i) $h_\alpha(u,v) > \dfrac{4\alpha}{(1-\alpha)\phi\,\mathrm{vol}\,(X)}$;

(ii) $\mathrm{vol}\,(X) \le \mathrm{vol}\,(G)/2$;

(iii) $\mathrm{vol}\,(Y)/\mathrm{vol}\,(X) \le (1+\phi/2)$.

Then $\phi_{uv} \le \phi$.

Proof. The proof is by contradiction. We first observe that $X \subset Y$. Suppose $\phi_{uv} > \phi$. Then we have

(15) $e(X,\bar{Y}) \ge e(X,\bar{X}) - \mathrm{vol}\,(Y\backslash X)$

$$\geq \phi \operatorname{vol}(X) - (\phi/2)\operatorname{vol}(X) \qquad \text{by hypothesis and (iii)}$$

$$\geq (\phi/2)\operatorname{vol}(X).$$

By (i) and Lemma 2, we have

$$\sum_{\substack{\{x,y\}\in E \\ x\in X, y\in \bar{X}}} h_\alpha(x,y) \geq \sum_{\substack{\{x,y\}\in E \\ x\in X, y\in \bar{Y}}} h_\alpha(x,y)$$

$$\geq e(X,\bar{Y}) h_\alpha(u,v)$$

since $h_\alpha(u,v)$ is the smallest value among all edges (x,y) such that $x \in X$ and $y \in \bar{Y}$. By inequality (15) and (i), the above inequality implies that

$$\sum_{\substack{\{x,y\}\in E \\ x\in X, y\in \bar{X}}} h_\alpha(x,y) > \frac{2\alpha}{1-\alpha}.$$

This contradicts Lemma 2 and the proof is complete. ■

7. Using PageRank to Estimate the Effective Resistance

The effective resistance $R(u,v)$ is a very useful graph invariant. For example, the recent work by Spielman and Srivastava [14] on graph sparsification relies on the effective resistance. Here we wish to illustrate that we can use the generalized effective resistance to approximate the effective resistance. Then by using the effective approximation algorithm for computing PageRank [1] and the PageRank representation in (11) for the generalized effective resistance, we can approximate the effective resistance as a result.

Recall that for $\alpha = \beta/(2+\beta)$, we have

$$R_\alpha(u,v) = \beta(\chi_u - \chi_v)D^{-1/2}\mathcal{G}_\beta D^{-1/2}(\chi_u - \chi_v)^T.$$

Theorem 9. *For two distinct vertices u and v in G, we have*

$$\left|\beta R(u,v) - R_\alpha(u,v)\right| \leq \frac{\beta^2}{\lambda_1^2}\left(\frac{1}{d_u} + \frac{1}{d_v}\right)$$

where λ_1 is the smallest nontrivial eigenvalue of $\mathcal{L}$ of G and $\alpha = \beta/(2+\beta)$.

Proof. From (3) and (8), we have

$$R(u,v) = (\chi_v - \chi_u)D^{-1/2}\mathcal{G}D^{-1/2}(\chi_v - \chi_u)^T$$
$$= \sum_{i=1}^{n-1} \frac{1}{\lambda_i}\left(\frac{\phi_i(u)}{\sqrt{d_u}} - \frac{\phi_i(v)}{\sqrt{d_v}}\right)^2.$$

From equation (5), we have

$$R_\alpha(u,v) = \sum_{i=0}^{n-1} \frac{\beta}{\lambda_i + \beta}\left(\frac{\phi_i(u)}{\sqrt{d_u}} - \frac{\phi_i(v)}{\sqrt{d_v}}\right)^2.$$

Combining the above two expressions, we have, for all $\beta > 0$,

$$\left|\beta R(u,v) - R_\alpha(u,v)\right|$$
$$\leq \left(\frac{\phi_0(u)}{\sqrt{d_v}} - \frac{\phi_0(v)}{\sqrt{d_v}}\right)^2 + \sum_{i=1}^{n-1}\left(\frac{\beta}{\lambda_i} - \frac{\beta}{\lambda_i + \beta}\right)\left(\frac{\phi_i(u)}{\sqrt{d_u}} - \frac{\phi_i(v)}{\sqrt{d_v}}\right)^2$$
$$= \sum_{i=1}^{n-1} \frac{\beta^2}{\lambda_i(\lambda_i + \beta)}\left(\frac{\phi_i(u)}{\sqrt{d_u}} - \frac{\phi_i(v)}{\sqrt{d_v}}\right)^2$$

since $\phi_0 = \mathbf{1}D^{1/2}/\sqrt{\mathrm{vol}\,(G)}$.

For two fixed vertices u and v, the vector f_u, defined by $f_u(i) = \phi_i(u)$, for $i = 0, \ldots, n-1$, is orthogonal to f_v. This implies

$$\sum_{i=1}^{n-1} \frac{\phi_i(u)\phi_i(v)}{\sqrt{d_u d_v}} = -\frac{\phi_0(u)\phi_0(v)}{\sqrt{d_u d_v}} = -\frac{1}{\mathrm{vol}\,(G)}$$

Thus we have

$$\left|\beta R(u,v) - R_\alpha(u,v)\right|$$
$$\leq \frac{\beta^2}{\lambda_1(\lambda_1 + \beta)}\left(\frac{2}{\mathrm{vol}\,(G)} + \sum_{i=1}^{n-1}\left(\left(\frac{\phi_i(u)}{\sqrt{d_u}}\right)^2 + \left(\frac{\phi_i(v)}{\sqrt{d_v}}\right)^2\right)\right)$$
$$\leq \frac{\beta^2}{\lambda_1^2}\left(\frac{2}{\mathrm{vol}\,(G)} + \frac{1}{d_u} - \frac{1}{\sqrt{d_u\,\mathrm{vol}\,(G)}} + \frac{1}{d_v} - \frac{1}{\sqrt{d_v\,\mathrm{vol}\,(G)}}\right)$$
$$\leq \frac{\beta^2}{\lambda_1^2}\left(\frac{1}{d_u} + \frac{1}{d_v}\right)$$

since f_u and f_v are orthonormal vectors. ■

REFERENCES

[1] R. Andersen, F. Chung and K. Lang, Local graph partitioning using pagerank vectors, *Proceedings of 47th Annual IEEE Symposium on Foundations of Computer Science* (2006), pp. 475–486.

[2] S. Brin and L. Page, The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems,* **30(1–7)** (1998), pp. 107–117.

[3] A. k. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky and P. Tiwari, The electrical resistance of a graph captures its commute time and cover time, *Proceedings of the 26th ACM Symposium on Theory of Computing* (1989), pp. 574–586.

[4] F. Chung, *Spectal Graph Theory,* AMS Publication, 1997.

[5] F. Chung and S.-T. Yau, Discrete Green's Functions. *Journal of Combinatorial Theory, Series A,* **91(1–2)** (2000), pp. 191–214.

[6] H. Haveliwala, Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search, *IEEE Trans. Knowl. Data Eng.,* **15(4)** (2003), pp. 784–796.

[7] Glen Jeh and Jennifer Widom, Scaling personalized web search. *Proceedings of the 12th World Wide Web Conference (WWW)* (2003), pp. 271–279.

[8] G. Green, An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism, Nottingham, 1828.

[9] S. Guattery and G. L. Miller, Graph embeddings and laplacian eigenvalues, *Journal on Matrix Analysis and Applications,* **21(3)** (2000), pp. 703–723.

[10] G. Kirchhoff, Über die Auflösung der Gleichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird, *Ann. Phys. chem.,* **72** (1847), 497–508.

[11] G. Kirchhoff, On the motion of electricity in conductors (Über die Bewegung der Elecktricität in Leitem), *Annalen der Physik,* Vol. **102** (1857), pp. 529.

[12] C. St. J. A. Nash-Williams, Random walks and electronic currents in networks, *Proceedings of the Cambridge Philosophical Society,* **55** (1959), pp. 181–194.

[13] L. Lovász, Random walks on graphs: A survey, Combinatorics, Paul Erdős is Eighty, **2** (1993), pp. 1–46.

[14] D. A. Spielman and N. Srivastava, Graph sparsification by effective resistances, *Proceedings of the 40th ACM Symposium on Theory of Computing* (2008), pp. 563–568.

Fan Chung and Wenbo Zhao

University of California,
San Diego,
La Jolla, CA 92093,
US

e-mail: {fan,w3zhao}@ucsd.edu

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 63–93.

SOLUTION OF PETER WINKLER'S PIZZA PROBLEM*†

JOSEF CIBULKA, JAN KYNČL, VIOLA MÉSZÁROS,
RUDOLF STOLAŘ and PAVEL VALTR

Bob cuts a pizza into slices of not necessarily equal size and shares it with Alice by alternately taking turns. One slice is taken in each turn. The first turn is Alice's. She may choose any of the slices. In all other turns only those slices can be chosen that have a neighbor slice already eaten. We prove a conjecture of Peter Winkler by showing that Alice has a strategy for obtaining 4/9 of the pizza. This is best possible, that is, there is a cutting and a strategy for Bob to get 5/9 of the pizza. We also give a characterization of Alice's best possible gain depending on the number of slices. For a given cutting of the pizza, we describe a linear time algorithm that computes Alice's strategy gaining at least 4/9 of the pizza and another algorithm that computes the optimal strategy for both players in any possible position of the game in quadratic time. We distinguish two types of turns, shifts and jumps. We prove that Alice can gain 4/9, 7/16 and 1/3 of the pizza if she is allowed to make at most two jumps, at most one jump and no jump, respectively, and the three constants are the best possible.

1. INTRODUCTION

Peter Winkler posed the following problem at the conference Building Bridges, honouring the 60th birthday of László Lovász, in Budapest in 2008. Bob and Alice are sharing a pizza. Bob cuts the pizza into slices

*Work on this paper was supported by the projects 1M0545 and MSM0021620838 of the Ministry of Education of the Czech Republic. Viola Mészáros was also partially supported by OTKA Grant K76099. Josef Cibulka and Rudolf Stolař were also supported by the Czech Science Foundation under the contract no. 201/09/H057.

†A short version of the paper appeared in *Combinatorial Algorithms,* the proceedings of the IWOCA 2009, Hradec nad Moravicí, Czech Republic, LNCS, Vol. 5874, pp. 356–368, *Springer* (2009).

of not necessarily equal size. Afterwards they take turns alternately to divide it among themselves. One slice is taken in each turn. In the first turn Alice takes any slice. In the forthcoming turns one may take a slice if it is adjacent to some previously taken slice. This is called the Polite Pizza Protocol. How much of the pizza can Alice gain?

The original puzzle, that is determining whether Bob can get more than half of the pizza, was devised by Dan Brown in 1996. Bob can easily ensure for himself one half of the pizza. For example, he may cut the pizza into an even number of slices of equal size. Then Bob always obtains exactly one half. Peter Winkler found out that Bob can actually get 5/9 of the pizza if he cuts the pizza properly – see Theorems 4 and 5 for such cuttings. He conjectured that Alice can obtain 4/9 of the pizza for any cutting. The main aim of this paper is to show a strategy of Alice proving this conjecture.

The pizza after Bob's cutting may be represented by a circular sequence $P = p_0 p_1 \dots p_{n-1}$ and by *the sizes* $|p_i| \geq 0$ (for $i = 0, 1, \dots, n-1$); for simplicity of notation, throughout the paper we do not separate the elements of (circular) sequences by commas. *The size of* P is defined by $|P| := \sum_{i=0}^{n-1} |p_i|$. Throughout the paper the indices are counted modulo n.

For $1 < j \leq n$, if one of the players chooses a slice p_i in the $(j-1)$-st turn and the other player chooses p_{i-1} or p_{i+1} in the j-th turn, then the j-th turn is called a *shift*, otherwise it is called a *jump*. Except for the first and the last turn, there are two choices in each turn and exactly one of them is a shift and the other one is a jump. The last turn is always a shift.

If some strategy of a player allows the player to make at most j jumps, then we call it a *j-jump strategy*. We remark that given a circular sequence P of length n, Alice has exactly n zero-jump strategies on P, determined by Alice's first turn.

Let Σ be a particular strategy of one of the players. We say that Σ is a strategy with *gain* g if it guarantees the player a subset of slices with the sum of sizes at least g. Note that according to this definition, if Σ is a strategy with gain g then it is also a strategy with gain g' for any $g' \leq g$.

If the number of slices is even, Alice has the following zero-jump strategy with gain $|P|/2$. She partitions the slices of the pizza into two classes, even and odd, according to their parity in P. In the first turn Alice takes a slice from the class with the sum of slice sizes at least $|P|/2$. In all her forthcoming turns she makes shifts, thus forcing Bob to eat from the other class in each of his turns.

Here is our main result.

Theorem 1. *For any P, Alice has a two-jump strategy with gain $4|P|/9$.*

More generally, we determine Alice's guaranteed gain for any given number of slices.

Theorem 2. *For $n \geq 1$, let $g(n)$ be the maximum $g \in [0,1]$ such that for any cutting of the pizza into n slices, Alice has a strategy with gain $g|P|$. Then*

$$g(n) = \begin{cases} 1 & \text{if } n = 1, \\ 4/9 & \text{if } n \in \{15, 17, 19, \dots\}, \\ 1/2 & \text{otherwise.} \end{cases}$$

Moreover, Alice has a zero-jump strategy with gain $g(n)|P|$ when n is even or $n \leq 7$, she has a one-jump strategy with gain $g(n)|P|$ for $n \in \{9, 11, 13\}$, and she has a two-jump strategy with gain $g(n)|P|$ for $n \in \{15, 17, 19, \dots\}$.

If we make a restriction on the number of Alice's jumps we get the following results.

Theorem 3. (a) *Alice has a zero-jump strategy with gain $|P|/3$ and the constant $1/3$ is the best possible.*

(b) *Alice has a one-jump strategy with gain $7|P|/16$ and the constant $7/16$ is the best possible.*

Due to Theorem 2, the following theorem describes all minimal cuttings for which Bob has a strategy with gain $5|P|/9$.

Theorem 4. *For any $\omega \in [0,1]$, Bob has a one-jump strategy with gain $5|P|/9$ if he cuts the pizza into 15 slices as follows: $P_\omega = 0010100(1+\omega)0(2-\omega)00202$. These cuttings describe, up to scaling, rotating and flipping the pizza upside-down, all the pizza cuttings into 15 slices for which Bob has a strategy with gain $5|P|/9$.*

For $\omega = 0$ or $\omega = 1$, the cutting in Theorem 4 has slices of only three different sizes $0, 1, 2$. If all the slices have the same size, then Alice always gets at least half of the pizza. But two different slice sizes are already enough to obtain a cutting with which Bob gets $5/9$ of the pizza.

Theorem 5. *Up to scaling, rotating and flipping the pizza upside-down, there is a unique pizza cutting into 21 slices of at most two different sizes for which Bob has a strategy with gain $5|P|/9$. The cutting is 001010010101001010101.*

In Subsection 7.1 we describe a linear-time algorithm for finding Alice's two-jump strategy with gain $g(n)|P|$ guaranteed by Theorem 2.

Theorem 6. *There is an algorithm that, given a cutting of the pizza with n slices, performs a precomputation in time $O(n)$. Then, during the game, the algorithm decides each of Alice's turns in time $O(1)$ in such a way that Alice makes at most two jumps and her gain is at least $g(n)|P|$.*

There is also a straightforward quadratic-time dynamic algorithm finding optimal strategies for each of the two players.

Claim 7. *There is an algorithm that, given a cutting of the pizza with n slices, computes an optimal strategy for each of the two players in time $O(n^2)$. The algorithm stores an optimal turn of the player on turn for all the $n^2 - n + 2$ possible positions of the game.*

We remark that, unlike in Theorem 1, the number of Alice's jumps in her optimal strategy cannot be bounded by a constant. In fact, it can be as large as $\lfloor n/2 \rfloor - 1$ for $n \geq 2$ (see Observation 28 in Subsection 7.2). A similar statement holds for the number of Bob's jumps in his optimal strategy.

The following question is still open.

Problem 1. Is there an algorithm that uses $o(n^2)$ time for some precomputations and then computes each optimal turn in constant time?

We remark that we even don't know if Alice's optimal first turn can be computed in time $o(n^2)$.

Independently of us and approximately at the same time, K. Knauer, P. Micek and T. Ueckerdt [3] also proved Theorem 1 and some related results.

The paper is organized as follows. Theorem 1 is proved in Section 2. Section 3 contains examples of cuttings showing that the constant 4/9 in Theorem 1 cannot be improved. Section 4 is devoted to the proof of Theorem 2. Theorems 4 and 5 are proved in Section 5. Section 6 contains the proof of Theorem 3. The algorithms from Theorem 6 and Claim 7 are described in Subsection 7.1 and Subsection 7.2, respectively.

In a follow-up paper [1] we discuss generalizations of the Pizza Problem.

2. The Lower Bound

When the number of slices is even, Alice can always gain at least $|P|/2$. Here we prove the lower bound on her gain when $n \geq 3$ is odd.

2.1. Preliminaries

If the number of slices is odd, instead of the circular sequence $P = p_0p_1 \dots p_{n-1}$ we will be working with the related circular sequence $V = v_0v_1 \dots v_{n-1} = p_0p_2 \dots p_{n-1}p_1p_3 \dots p_{n-2}$ that we call *the characteristic cycle* (see Figure 1). The size of the characteristic cycle is denoted by $|V|$. Clearly $|V| = |P|$.

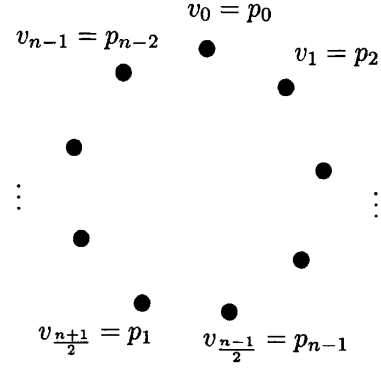


Fig. 1. A cutting of a pizza and the corresponding characteristic cycle

An *arc* is a sequence of at most $n-1$ consecutive elements of V. If we talk about the first or the last element of an arc, we always consider it with respect to the linear order on the arc inherited from the characteristic cycle V. For an arc $X = v_iv_{i+1} \dots v_{i+l-1}$, its *length* is $l(X) := l$ and its size is $|X| := \sum_{j=i}^{i+l-1} |v_j|$. An arc of length $(n+1)/2$ is called a *half-circle.* Figure 2 shows an example of a game on V. The slice taken in the i-th turn is labelled by the initial letter of the player with i in the subscript.

At any time during a game, a player may decide to make only shifts further on. The player will take one or two arcs of the characteristic cycle afterwards. An example of such a game when Alice decided to make no more jumps is depicted on Figure 3 (slices taken before the decision point are labelled with $*$, and selected pairs of slices neighboring in the original pizza are connected by dashed segments). The slices she took after the

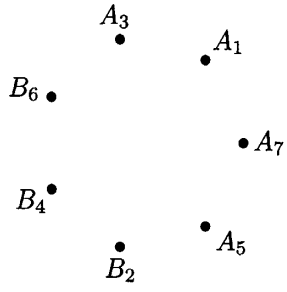


Fig. 2. A game illustrated on the characteristic cycle V (the turns are $A_1, B_2, A_3, \dots$). The turns B_4 and A_5 are jumps and all the other turns (except A_1) are shifts

decision point are forming two arcs that are separated in between by some arc of previously taken slices.

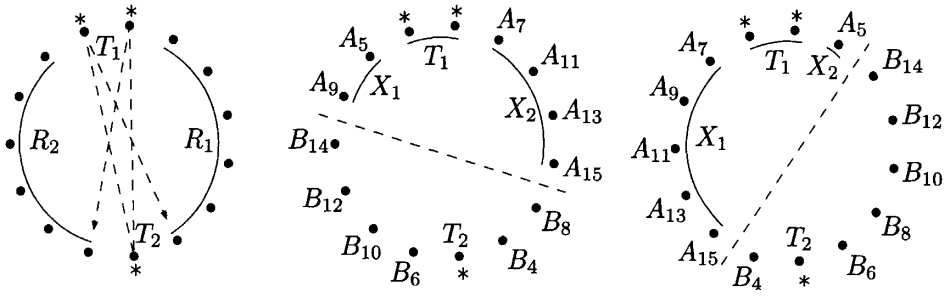


Fig. 3. Situation before Bobs turn with the two possible options marked by arrows (left) and two of the possible ends of the game where Alice made no more jumps (middle and right)

Observation 8. *Consider a position after Alice's turn $A_j, j \neq 1, n$. We have $V = T_1R_1T_2R_2$, where $\ell(T_1) = \ell(T_2) + 1 = (j+1)/2$, $\ell(R_1) = \ell(R_2)$, T_1 and T_2 are two arcs of already taken slices, and R_1 and R_2 are two arcs containing the remaining slices. Suppose that all the remaining turns of Alice $(A_{j+2}, A_{j+4}, \dots, A_n)$ are shifts. Then, regardless of Bob's remaining turns $B_{j+1}, \dots, B_{n-1}$, the slices taken by Alice in the turns $A_{j+2}, A_{j+4}, \dots, A_n$ necessarily form two arcs X_1 and X_2 such that $X_1T_1X_2$ is a half-circle of V.*

In addition, for any half-circle $Y_1T_1Y_2$, Bob can choose his turns B_{j+2}, $\dots, B_{n-1}$ so that $X_1 = Y_1$ and $X_2 = Y_2$.

Proof. We will show by induction that before any Bob's turn B_{j+2k+1}, the slices taken by him in turns $B_{j+1}, \dots, B_{j+2k-1}$ form two arcs Z_1 and Z_2 such that $Z = Z_1T_2Z_2$ is an arc and his two possible moves are on the

two neighbors of Z. This is true for B_{j+1} and by induction if this is true before B_{j+2k+1}, then Bob takes for B_{j+2k+1} one of the two neighbors of Z and $Z' := Z \cup B_{j+2k+1}$ is an arc. After Alice's shift, Bob's shift would be a neighbor of B_{j+2k+1}, thus a neighbor of Z'. Bob's jump would be the neighbor of Z different from B_{j+2k+1}, thus a neighbor of Z'.

For any given half-circle $Y_1T_1Y_2$ and before any of Bob's turns $B_{j+1}, \ldots, B_{n-1}$, the two slices available for Bob are neighbors of an arc of length at most $(n-3)/2$ which is not a subarc of $Y_1T_1Y_2$. Thus one of the two slices available for him is not in $Y_1T_1Y_2$ and Bob can choose his turns $B_{j+1}, \ldots, B_{n-1}$ so that $X_1 = Y_1$ and $X_2 = Y_2$. ■

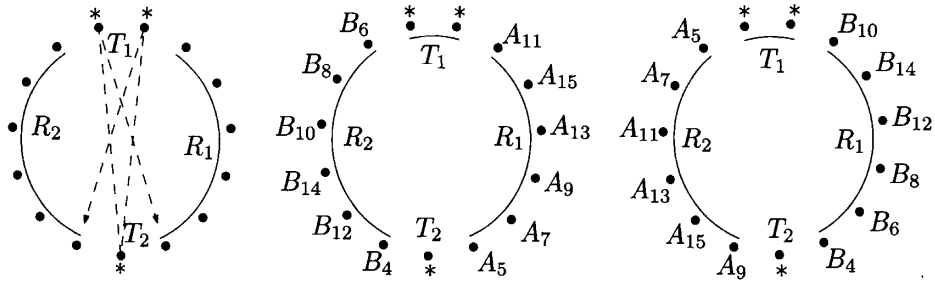


Fig. 4. Two possible choices of Bob's next turn (left) and the two possible ends of the game where Bob made no more jumps (middle and right)

If Bob decides to make only shifts for the rest of the game, he takes one arc afterwards. Namely, if there are two arcs of already taken slices in V at his decision point, then the arc that will be taken by Bob is neighboring these two arcs at both of its ends (see Figure 4).

Observation 9. *Consider a position after Alice's turn $A_j, j \neq 1, n$. We have $V = T_1R_1T_2R_2$, where $\ell(T_1) = \ell(T_2) + 1 = (j+1)/2$, $\ell(R_1) = \ell(R_2)$, T_1 and T_2 are two arcs of already taken slices, and R_1 and R_2 are two arcs containing the remaining slices. Bob's turn B_{j+1} may be on the last slice of R_1 or on the first slice of R_2. If B_{j+1} is on the last slice of R_1 and all the remaining turns of Bob are shifts then, regardless of Alice's remaining turns, Bob will take R_1 and Alice will take R_2 in this phase of the game. Similarly, if B_{j+1} is on the first slice of R_2 and all the remaining turns of Bob are shifts then, regardless of Alice's remaining turns, Bob will take R_2 and Alice will take R_1 in this phase of the game.*

Proof. Similarly to the proof of Observation 8, it is easy to prove by induction that if Bob played B_{j+1} on R_1, then before each Alice's turn, the two slices available for her are from R_2. ■

2.2. Minimal triples

For each v in V *the potential of* v is the minimum of the sizes of half-circles covering v. The maximum of the potentials in V is *the potential of* V, which we further denote by $p(V)$. It is an immediate conclusion that Alice has a strategy with gain $p(V)$ because by choosing an element with potential equal to $p(V)$ and making only shifts afterwards Alice obtains at least $p(V)$. Therefore we may assume that $p(V) < |V|/2$.

A *covering triple* of half-circles is a triple of half-circles such that each element of V appears in at least one of the three half-circles. We allow two half-cirles to be equal in the covering triple. A covering triple is *minimal* if it contains a half-circle of minimum size (among all n half-circles), all half-circles forming the triple have size at most $p(V)$ and none of them may be replaced in the triple by a half-circle of strictly smaller size.

Claim 10. *Each half-circle of minimum size lies in at least one minimal triple.*

Proof. Take a half-circle H_1 of minimum size. Consider v_k and $v_{k+(n-3)/2}$ the two uncovered elements neighboring H_1 on V. Let H_2 be the half-circle of size at most $p(V)$ that covers v_k and as many elements of V not covered by H_1 as possible. We define H_3 in the same way for $v_{k+(n-3)/2}$. The above triple of half-circles covers V. If it is not the case, then take an uncovered element v. Consider a half-circle H that has minimal size among half-circles covering v. At least one of v_k and $v_{k+(n-3)/2}$ is covered by H. This contradicts the choice of H_2 or H_3. So we get that the given triple of half-circles forms a covering triple. Now while any of the half-circles can be replaced in the triple by a half-circle of strictly smaller size, we replace it. Obviously H_1 won't be replaced as it is a half-circle of minimum size. Consequently the triple we get is a minimal triple. ■

Observation 11. *If the size of a half-circle in a minimal triple is z then Alice has a zero-jump strategy with gain z.*

Proof. As in a minimal triple all half-circles are of size at most $p(V)$ and Alice has a zero-jump strategy with gain $p(V)$, the statement of the observation follows. ■

Claim 12. *Let $p(V) < |V|/2$. Then any minimal triple contains three pairwise different half-circles, and thus there is a partition of V into six*

arcs A, B, C, D, E, F such that the half-circles in the minimal triple are ABC, CDE and EFA (see Figure 5). The lengths of the arcs satisfy $l(A) = l(D) + 1 \geq 2$, $l(C) = l(F) + 1 \geq 2$ and $l(E) = l(B) + 1 \geq 2$.

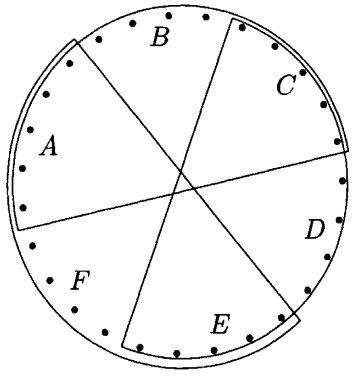


Fig. 5. The partitioning of the characteristic cycle given by the covering half-circles

Proof. If two of the three half-circles in a minimal triple are equal then V can be covered by two half-circles of the triple. Since each half-circle in the triple has size at most $p(V)$, the total size of the pizza is at most $2p(V) < |V|$, a contradiction. If at least one of B, D, F has length 0, we argue exactly in the same way.

We have $l(ABC) + l(EFA) = n + 1 = l(A) + \cdots + l(F) + 1$, therefore $l(A) = l(D) + 1 \geq 2$. The other two equalities are analogous. ∎

2.3. An auxiliary one-jump strategy

Throughout this section we assume that $p(V) < |V|/2$. We fix any minimal triple T of half-circles. By Claim 12, it yields a partition of V into six arcs A, B, C, D, E, F such that the half-circles in the triple are ABC, CDE, EFA (see Figure 5). We further use the notation $a := |A|$, $b := |B|$, and so on.

We define a *median slice* of an arc $X = v_i v_{i+1} \ldots v_{i+l}$ to be a slice $v_k \in X$ such that $\sum_{j=i}^{k-1} |v_j| \leq |X|/2$ and $\sum_{j=k+1}^{i+l} |v_j| \leq |X|/2$. Observe that any arc of positive length has at least one median slice.

Claim 13. *Alice has a one-jump strategy for V with gain $b/2 + \min\{c + d, f + a\}$ if $p(V) < |V|/2$.*

Proof. By Claim 12 we have that $l(B) > 0$. In the first turn Alice takes a median slice v_k of B. Consequently Bob is forced to start in E. He may take the element $v_{k+(n-1)/2}$ or $v_{k+(n+1)/2}$. Alice makes only shifts while the shift implies taking an element of B. In the meantime Bob necessarily takes elements from E. In the turn, when Alice's shift would imply taking an element outside of B, Alice makes a jump instead. In that moment some initial arc E_0 of E starting from the boundary of E is already taken. Let E_1 be the remaining part (subarc) of E. Alice takes the available element of E_1. Note that such an element exists since in the sequence P, all the neighbors of the slices of B are in E. (see Figure 6). The half-circle $X_1E_0X_2$ can replace either CDE or EFA in the fixed minimal triple. Thus due to the minimality of the triple, the size of X_1X_2 is always at least the size of either CD or FA. As Alice obtained at least the half of B before the jump, in the end she gains at least $b/2 + \min\{c+d, f+a\}$. ■

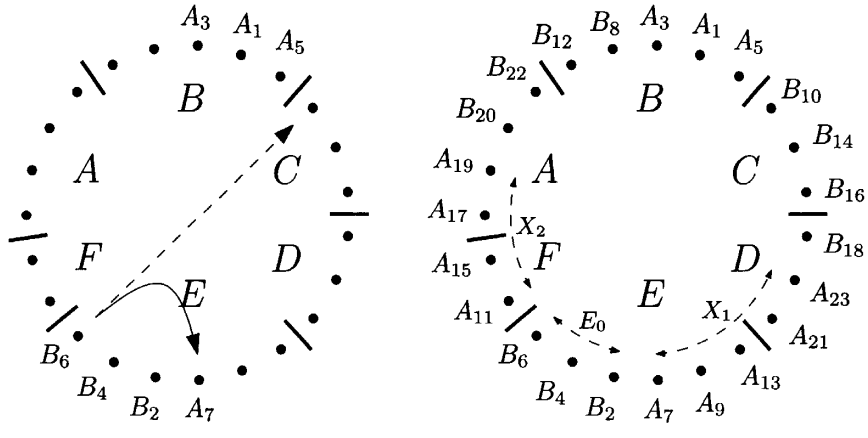


Fig. 6. One-jump strategy: Alice chooses a jump rather than a shift (left) and makes no more jumps afterwards (right)

Corollary 14. *Alice has a one-jump strategy for V with gain $(a+b+c)/4+(d+e+f)/2$ if $p(V) < |V|/2$.*

Proof. By Claim 13 Alice has a strategy with gain $b/2+\min\{c+d, f+a\}$. Without loss of generality we may assume this sum is $g_1 := b/2 + c + d$. Alice also has a strategy with gain $g_2 := e + f + a$ by Observation 11. Combining the two results Alice has a gain $\max\{g_1, g_2\} \geq g_1/2 + g_2/2 = (a+c+d+e+f)/2 + b/4 \geq (a+b+c)/4 + (d+e+f)/2$. ■

2.4. A two-jump strategy

Throughout this subsection we assume that $p(V) < |V|/2$ and that V is partitioned into six arcs $A, \ldots, F$ in the same way as in the previous subsection.

In this subsection we describe a strategy satisfying the following claim.

Claim 15. *Alice has a two-jump strategy for V with gain $b/2 + e/4 + \min\{c + d, f + a\}$ if $p(V) < |V|/2$.*

2.4.1. Two phases of the game. Let $B = v_i v_{i+1} \ldots v_{i+\Delta}$. Then $E = v_j v_{j+1} \ldots v_{j+\Delta+1}$, where $j = i + (n-1)/2$. Consider the circular sequence $V' = v_i v_{i+1} \ldots v_{i+\Delta}\, v_j\, v_{j+1} \ldots v_{j+\Delta+1}$ obtained by concatenating the arcs B and E.

Let H be a half-circle of V' containing v_j. Then its size is not smaller than the size of E, since otherwise the half-circle CDE of V could be replaced in the minimal triple T by a half-circle of smaller size – namely by the half-circle formed by the slices contained in CD and in H.

Similarly, if H is a half-circle of V' containing $v_{j+\Delta+1}$, then its size is also not smaller than the size of E. Since each half-circle of V' contains v_j or $v_{j+\Delta+1}$, it follows that E is a half-circle of V' of minimum size.

If $p(V') \geq |V'|/2$ then Alice has a zero-jump strategy Σ for V' with gain $p(V') \geq |V'|/2 \geq b/2 + e/4$. Otherwise, by Corollary 14 (applied on V'), Alice has a one-jump strategy Σ for V' with gain $b/2 + e/4$ (we use the fact that E is a half-circle of V' of minimum size, and therefore it is contained in a minimal triple yielding a partition of V' into six arcs $A', B', \ldots, F'$ such that $E = A'B'C'$ and $B = D'E'F'$).

Briefly speaking, Alice's strategy on V follows the strategy Σ as long as it is possible, then Alice makes one jump and after that she makes only shifts till the end of the game.

In the rest of this subsection (Subsection 2.4), we consider a game G on V. We divide the turns of G into two phases. *The first phase of G* is the phase when Alice follows the strategy Σ and it ends with Bob's turn. Alice's first turn that does not follow (and actually cannot follow) the strategy Σ is the first turn of *the second phase of G*. It is always a jump and all the other turns of Alice in the second phase are shifts.

We now describe Alice's strategy in each of the two phases of G in detail.

2.4.2. Alice's strategy in the first phase. As mentioned above, Alice has a one-jump strategy Σ for V' with gain $b/2 + e/4$. We now distinguish two cases.

Case 1: The strategy Σ is a zero-jump strategy. Let the first turn in the zero-jump strategy Σ be on a slice $q \in V'$. The first turn could be also on any other point of V' with the same or larger potential. Observe that the potentials of the slices in V' are e on E and at least e on B. Therefore we may assume that q lies in B.

In the game G, Alice makes her first turn also on q. In the second turn Bob can choose between two slices in E. In the subsequent turns Alice makes shifts as long as Bob's previous turn was neither on the first nor on the last slice of E. Consider all slices taken by Bob up to any fixed moment during the first phase of the game G. They always form a subarc of E (and the slices taken by Alice form a subarc of B). The first turn in which Bob takes the first or the last slice of E is the last turn of the first phase. Note that after that Alice's shift would be either on the last slice of A or on the first slice of C (see Figure 7). But Alice makes a jump and this jump is the first turn of the second phase. Note that this jump is in E (see Figure 7).

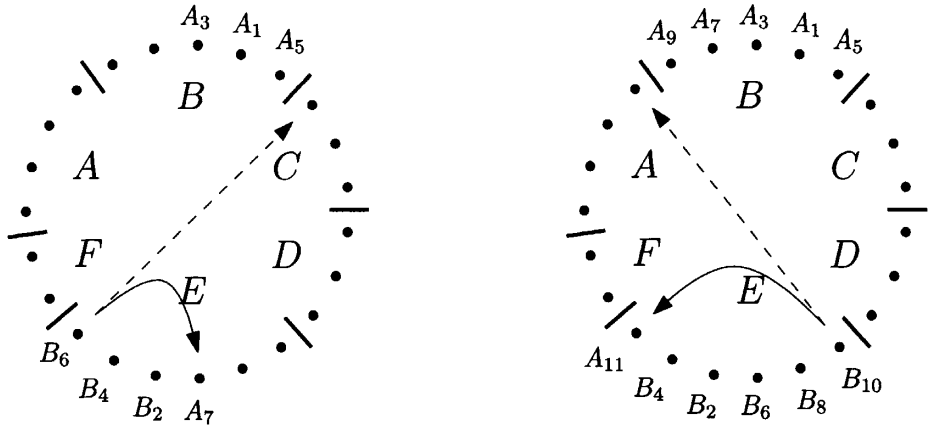


Fig. 7. Case 1: After the end of first phase, Alice chooses a jump rather than a shift (two examples shown)

Case 2: The strategy Σ is not a zero-jump strategy. Following the proof of Corollary 14, we may suppose that Σ is the strategy that we describe below.

By Claim 10, the half-circle E of minimum size is contained in some minimal triple T' of half-circles of V'. The triple T' determines a partition

of V' into six arcs $A', B', \ldots, F'$ in the same way as T determined a partition of V into $A, B, \ldots, F$. We may suppose that $E = A'B'C'$ and $B = D'E'F'$.

We may suppose that the size of B' is positive, since otherwise one of the half-circles $C'D'E'$ and $E'F'A'$ has size at least $b/2+e/2$ and thus Alice has a zero-jump strategy for V' with gain $b/2+e/2$, allowing us to use the above Case 1.

In the first turn Alice takes a median slice of B'. Then in the second turn Bob can choose between two slices of E'. In the subsequent turns Alice makes shifts as long as Bob's previous turn was neither on the first nor on the last slice of E'. In each moment in this part of the game Bob's turns form a subarc of E'. At the first instance when Bob takes the first or the last slice of E', Alice makes a jump, which is always in E' (see Figure 8). Note that so far the game was an analogue of the first phase in Case 1, with B' and E' in place of B and E, respectively. After her first jump Alice makes shifts as long as Bob's previous turn was neither on the first nor on the last slice of E. Note that Bob's turns in this part of the game are in E (see Figure 8). At the first instance when Bob takes the first or the last slice of E, Alice makes a jump, which is already the first turn of the second phase. This jump is necessarily in E (see Figure 8).

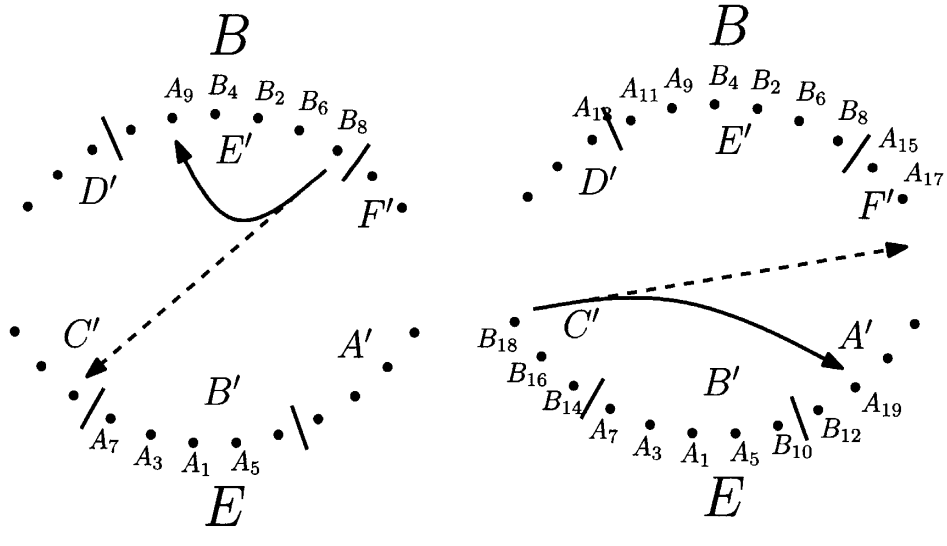


Fig. 8. Case 2: During the first phase, Alice makes a jump rather than a shift (left) and then she makes an other jump after the end of the first phase (right)

2.4.3. Alice's strategy in the second phase. Alice's strategy in the second phase is very simple. Above we describe the first phase and also the first turn of the second phase, which is always a jump done by Alice. In the rest of the second phase Alice makes only shifts.

2.4.4. Analysis of Alice's gain. Since the first phase of G ends by Bob's turn on the first or on the last slice of E, we may suppose without loss of generality that it ends with Bob's turn on v_j. Then the part of V removed in the first phase of G is a union of some initial subarc B_0 of B and some initial subarc E_0 of E. Let E_1 be the arc formed by the slices of E not taken in the first phase of G, thus $E = E_0E_1$, and let $e_1 := |E_1|$. In her jump at the beginning of the second phase of G, Alice takes the first slice of E_1.

By Observation 8, all the slices taken by Alice in the second phase of G form two arcs X_1 and X_2 such that $X_1E_0X_2$ is a half-circle of V (see Figure 6). Since none of the half-circles CDE and EFA can be replaced in the triple T by a half-circle of a strictly smaller size, the sum $|X_1| + |X_2|$ achieves its minimum either for $X_1 = CD$ and $X_2 = E_1$, or for $l(X_1) = 0$ and $X_2 = E_1FA$. Thus, the portion collected by Alice in the second phase of G is at least $e_1 + \min\{c + d, f + a\}$.

Now, consider an auxiliary game G' on V' consisting of two phases defined as follows. The turns in *the first phase of* G' are exactly the same as the turns in the first phase of G (this is a correct definition, as all turns in the first phase of G are in $B \cup E$ and the first or the last slice of E is taken only at the very end of the first phase). In *the second phase of* G', both Alice and Bob make only shifts.

We claim that Alice actually follows the one-jump strategy Σ in the whole game G'. This is obvious in the first phase of G'. Further, if Σ is a zero-jump strategy then Alice clearly follows Σ also in the second phase of the game G'. Otherwise Alice makes her only jump in the first phase of the game (see Case 2 in Paragraph 2.4.2) and thus again she follows Σ also in the second phase of G'.

By Observation 9, Alice collects exactly the slices of E_1 in the second phase of G'. Thus, if g denotes the portion collected by Alice in the first phase of G' then $g + e_1$ is her portion in the whole game G'. Since the strategy Σ guarantees gain $b/2 + e/4$ to Alice, we get $g + e_1 \geq b/2 + e/4$.

Alice's portion in the whole game G is at least $g + \big(e_1 + \min\{c + d, f + a\}\big) \geq b/2 + e/4 + \min\{c + d, f + a\}$, which completes the proof of Claim 15.

2.5. Proof of the lower bound

Proof of Theorem 1. If the number of slices is even then Alice has a zero-jump strategy with gain $|P|/2$.

We further suppose that the number of slices is odd. We consider the characteristic circle V. If $p(V) \geq |V|/2$ then Alice has a zero-jump strategy with gain $|V|/2 = |P|/2$.

Suppose now that $p(V) < |V|/2$. Then V may be partitioned into six arcs $A, \ldots, F$ as in Claim 12. Without loss of generality, we may assume that $a+b+c \leq c+d+e \leq e+f+a$. Thus, $a+b \leq d+e$ and $c+d \leq f+a$. By Observation 11, Alice has a zero-jump strategy with gain

$$g_1 := e + f + a.$$

By Claim 15, Alice has a two-jump strategy with gain

$$g_2 := b/2 + e/4 + \min\{c + d, f + a\} = b/2 + e/4 + c + d.$$

By an analogue of Claim 15, Alice also has a two-jump strategy with gain

$$g_3 := f/2 + c/4 + \min\{a + b, d + e\} = f/2 + c/4 + a + b.$$

One of the three strategies gives gain

$$\max\{g_1, g_2, g_3\} \geq (3g_1 + 4g_2 + 2g_3)/9$$

$$= (5a + 4b + 9c/2 + 4d + 4e + 4f)/9 = \left(4|P| + a + c/2\right)/9 \geq 4|P|/9. \quad \blacksquare$$

3. The Upper Bound

In this section we show a strategy for Bob that guarantees him 5/9 of the pizza. Then Bob has to cut the pizza into an odd number of slices, since otherwise Alice has a strategy with gain $|P|/2$, as was observed in the introduction. Before each turn of Bob, the number of the remaining slices is even. The sequence of all the remaining slices can be then written as

$$p_i p_{i+1} p_{i+2} \cdots p_{i+2j-1}.$$

Let $K := p_i p_{i+2} \ldots p_{i+2j-2}$ and $L := p_{i+1} p_{i+3} \ldots p_{i+2j-1}$ be the sequences of slices on odd and even positions respectively.

We use the following reformulation of Observation 9:

Observation 16. *Before any of his turns, Bob has a strategy that guarantees him* $\max\{|K|, |L|\}$ *in addition to what he already has. In the strategy Bob makes only shifts, except possibly for the first turn.*

Proof. We prove that Bob can get all slices from L. A similar proof shows that he can get all slices from K. In his first turn, Bob takes $p_{i+2j-1} \in L$. In each other turn, Bob makes shifts. Then before each of Alice's turns, the two slices available for her are from K. ■

Claim 17. *Bob has a one-jump strategy with gain* $5|P|/9$ *if he cuts the pizza into* 15 *slices in the following way:* 002020030300404.

Proof. The size of the pizza is 18, which means that Bob wants to get slices with sum of sizes at least 10.

We consider all possible first moves of Alice:

1. Alice takes a zero slice located between two nonzero slices. The sizes of the slices remaining after her turn are

   $$\underline{a}0\underline{0}\underline{b}0\underline{b}\underline{0}0\underline{c}0\underline{c}\underline{0}0\underline{a},$$

   where the elements of K are underlined and a, b and c are in one of the six possible bijections with 2, 3 and 4. Then $|K| = 2c + a$ and $|L| = 2b + a$ and by a case analysis of the possible values of a, b and c, Bob gets $\max\{|K|, |L|\} \geq 10$.

2. Alice takes a zero slice located between a zero slice and a nonzero slice. This leads to

   $$\underline{a}0\underline{a}\underline{0}0\underline{b}0\underline{b}\underline{0}0\underline{c}0\underline{c}0$$

   and Bob gets $\max\{|K|, |L|\} = \max\{2a + 2c, 2b\} = 2a + 2c \geq 10$.

3. Alice takes a nonzero slice. The situation is then

   $$\underline{0}\underline{a}\underline{0}0\underline{b}0\underline{b}\underline{0}0\underline{c}0\underline{c}\underline{0}0.$$

   Bob now takes the rightmost slice and then makes shifts until he either takes a or the two slices c. This leads to three possible cases:

   $$(a)\ \ \underline{0}0\underline{b}0\underline{b}\underline{0}0\underline{c}0\underline{c}\underline{0},$$

   $$(b)\ \ \underline{0}0\underline{b}0\underline{b}\underline{0}0\underline{c}\underline{0},$$

(c) $\underline{0}a\underline{00}b\underline{0}b\underline{00}$.

After Alice takes one of the available zero slices, we use Observation 16 to show that the gain of Bob in these three cases is

(a) $a + \max\{|K|, |L|\} = a + \max\{2b, 2c\}$,

(b) $a + c + \max\{|K|, |L|\} = a + c + \max\{2b, c\} = a + c + 2b$,

(c) $2c + \max\{|K|, |L|\} = 2c + \max\{2b, a\} = 2c + 2b$.

In any of the three cases and for any bijective assignment of the values 2, 3 and 4 to a, b and c, Bob gets slices of total size at least 10. ■

Corollary 18. *For any $\omega \in [0, 1]$, Bob has a one-jump strategy with gain $5|P|/9$ if he cuts the pizza into 15 slices as follows: $P_\omega = 0010100(1+\omega)0(2-\omega)00202$.*

Proof. The claim holds for $\omega = 1/2$, since $P_{1/2}$ is a scale-down of the pizza considered in Claim 17.

Clearly, if some slices of $P_{1/2}$ have total size at least 5 then also the corresponding slices of $P_\omega, \omega \in [0, 1]$, have total size at least 5. Therefore, Bob can ensure gain $5|P_\omega|/9 = 5$ for $P_\omega, \omega \in [0, 1]$, with the same strategy as for $P_{1/2}$. ■

In Section 4 we show that Bob has no strategy with gain $5|P|/9$ for pizza cuttings with fewer than 15 slices. Moreover, in Section 5 we show that Corollary 18 describes essentially all cuttings into 15 slices that guarantee Bob 5/9 of the pizza.

Claim 19. *For any odd $n \geq 15$, Bob has a one-jump strategy with gain $5|P|/9$ using some cutting of the pizza into n slices.*

Proof. For $n = 15$ the claim follows from Claim 17. For larger n, we take the cutting P from Claim 17 and add $n - 15$ zero slices between some two consecutive zero slices in the sequence.

If Alice starts in one of the added slices, then the situation is similar to the case 2 of the proof of Claim 17. The only difference is that there might be additional zeros at the beginning and at the end of the sequence. But these zeros either do not change the values of $|K|$ and $|L|$ or swap the two values. Thus Bob can get $\max\{|K|, |L|\} = 10$.

Otherwise Bob uses the strategy from the proof of Claim 17. In cases 1, 2, the even number of consecutive newly added zero slices does not change the value of $\max\{|K|, |L|\}$ and Bob can thus get slices of total size 10. In case 3, Alice first takes the slice of size a. The even number of added zero slices does not change the fact that before she is able to take any other nonzero slice, Bob takes either the slice of size a or the two slices of size c. After this, the value of $\max\{|K|, |L|\}$ is the same as in the proof of Claim 17. ■

For $\omega = 0$ or $\omega = 1$, the cutting used in Corollary 18 has slices of only three different sizes $0, 1, 2$. If all the slices have the same size, then Alice always gets at least half of the pizza. But two different slice sizes are already enough to obtain a cutting with which Bob gets $5/9$ of the pizza:

Claim 20. *If Bob can make slices of only two different sizes, then he can gain* $5/9$ *of the pizza by cutting the pizza into* 21 *slices of sizes* 0 *and* 1 *in the following way:* 001010010101001010101.

Moreover for any odd $n \geq 21$, *there is a cutting with* n *slices of two different sizes of slices, for which Bob has a one-jump strategy with gain* $5|P|/9$.

Proof. Let $n \geq 21$ and let the cutting be $(00)^m 1010101001010010101$ where $m \geq 1$. After Alice makes her first turn, there will be an even number of slices remaining and they can be partitioned into two classes based on their parity in the new sequence. If Alice takes a slice of size 0 in the first turn, necessarily one of the classes will contain more slices of size 1 than the other class. Bob's strategy will be to take the available element of the bigger class in each turn. Consequently Alice will be forced to take an element of the smaller class in each turn. If in the first turn Alice takes a slice of size 1 but the previous partitioning still gives two classes of different size, Bob uses the same strategy. Obviously Bob gains at least $5/9$ of the pizza in these cases.

There are three slices of size 1 indicated by bold which if taken in the first turn result in a partitioning into two classes of equal size:

$$(00)^m 1010\mathbf{1}010010\mathbf{1}0010\mathbf{1}0101.$$

The sequences below show the situation after the first turn, the vertical line indicates the place of the taken slice:

1. $(00)^m$ 1010|0100101 0010101

2. $(00)^m 101010100\,\boxed{10|\underline{0}0010101}$

3. $(00)^m 101010100\,\boxed{10100|\underline{0}101}$

Bob takes the underlined slice of size 0 in his first turn. He will make shifts until he gets to one of the borders of the boxes indicated above, then he makes a jump in his next turn and afterwards only shifts till the end of the game.

Regardless of what turns Alice makes, depending on the case Bob will obtain at least one or two slices of size 1 before anybody reaches outside the given box. As he makes a jump in his next turn and later only shifts, he will get all slices of size 1 outside the box. Therefore in all cases his gain adds up to at least 5/9 of the pizza.

The claim can also be checked for any fixed odd number $n \geq 21$ with a computer program based on the algorithm from Section 7.2. ■

4. Fixed Number of Slices

Here we prove Theorem 2. The theorem is trivial for $n = 1$ and easy for n even as observed in the introduction. Further, the theorem for $n \in \{15, 17, 19, \dots\}$ follows from Theorem 1 and Claim 17. An upper bound $g(n) \leq 1/2$ for any $n \geq 2$ can be seen on the pizza $1100\dots00$. It remains to show that Alice has a one-jump strategy with gain $1/2$ for any pizza $p_1 p_2 \dots p_n$, where n is odd and $3 \leq n \leq 13$.

Let n be odd and let $3 \leq n \leq 13$. We partition the characteristic cycle $V = v_1 v_2 \dots v_n$ into six arcs $A, B, \dots, E$ in the same way as in Section 2. We may suppose that each of the six arcs has a positive length, since otherwise Alice has a zero-jump strategy with gain $|P|/2$ (by Claim 12). Therefore, as $l(A) = l(D) + 1, l(C) = l(F) + 1$ and $l(E) = l(B) + 1$, and $n \leq 13$, at least one of the arcs B, D, F has length at most 1 (and hence, exactly 1). Due to the symmetries, it therefore suffices to prove the following claim:

Claim 21. *If $l(B) = 1$ then Alice has a one-jump strategy with gain $1/2$.*

Proof. First we describe a one-jump strategy with gain $b + \min\{c+d, f+a\}$. Alice's first turn in the strategy is on the only slice of B. Recall that $l(E) = l(B) + 1 = 2$. Bob can choose between the two slices of E in the

second turn. Alice takes the other slice of E in the third turn of the game. In the rest of the game, Alice makes only shifts, thus collecting slices of some pair of arcs X_1 and X_2 such that X_1EX_2 is a half-circle (see Figure 6). Since none of the half-circles CDE and EFA can be replaced in the triple T by a half-circle of a strictly smaller size, the sum $|X_1| + |X_2|$ achieves its minimum either for $X_1 = CD$ and $X_2 = \emptyset$, or for $X_1 = \emptyset$ and $X_2 = FA$. Thus, the portion collected by Alice in the whole game is at least

$$g_1 := b + \min\{c + d, f + a\}.$$

By two applications of Observation 11, Alice also has a zero-jump strategy with gain

$$g_2 := \max\{c + d + e, e + f + a\}.$$

One of the two strategies is a one-jump strategy with gain

$$\max\{g_1, g_2\} \geq (g_1 + g_2)/2 = 1/2. \qquad \blacksquare$$

5. Cuttings into 15 and 21 Slices

Here we prove that Bob's cuttings described in Section 3 include all pizza cuttings into 15 and into at most 21 slices where he gets his best possible gain. Theorem 2 implies the minimality of 15 slices as well.

Claim 22. *Corollary 18 describes, up to scaling, rotating and flipping the pizza upside-down, all the pizza cuttings into* 15 *slices for which Bob has a strategy with gain* $5|P|/9$.

Proof. Suppose Bob cuts the pizza into 15 slices so that Alice cannot gain more than $4|P|/9$. It follows that $p(V) \leq 4/9 < 1/2$. By Claim 12, the characteristic cycle can be partitioned into non-empty arcs A, B, C, D, E, F such that ABC, CDE and EFA form a minimal triple.

Following the proof of Theorem 1, we assume without loss of generality that $a+b+c \leq c+d+e \leq e+f+a$ and we consider the same three strategies with gains g_1, g_2 and g_3. Combining the assumption $g_1, g_2, g_3 \leq 4|P|/9$ with the inequalities $(3g_1 + 4g_2 + 2g_3)/9 \geq (4|P| + a + c/2)/9 \geq 4|P|/9$ we get equalities everywhere. Consequently, $g_1 = g_2 = g_3 = 4|P|/9$ and $a = c = 0$. This, in particular, implies that $e + f = b/2 + e/4 + d = f/2 + b = 4|P|/9$.

Now we show that $e = 0$. Applying Claim 21 three times, we get that $l(B) = l(D) = l(F) = 2$ and $l(A) = l(C) = l(E) = 3$. Particularly, the length of the circular sequence V' obtained by concatenating arcs B and E is 5. Hence, by Theorem 2, Alice has a strategy with gain $b/2 + e/2$ on V'. Following the proof of Claim 15 we get a two-jump strategy on V with gain $b/2 + e/2 + c + d = g_2 + e/4 = 4|P|/9 + e/4$. Therefore $e = 0$.

It follows that $f = 4|P|/9, b = 2|P|/9$ and $d = 3|P|/9$. If one of the two slices in B had size greater than $|P|/9$, then by the proof of Claim 13 Alice would have a one-jump strategy with gain greater than $|P|/9 + c + d = 4|P|/9$. It follows that both slices in B have size exactly $|P|/9$. Similarly, using an analogue of Claim 13, we conclude that both slices in F have size exactly $2|P|/9$ and both slices in D have size at most $2|P|/9$. ■

Lemma 23. *Let $P = p_1p_2 \dots p_n$ be a cutting of a pizza into an odd number of slices for which Bob has a strategy with gain $g \geq |P|/2$. Let $x = \min_{i\in\{1,2,\dots,n\}} |p_i|$ and let $P' = p'_1p'_2 \dots p'_n$ be a cutting of a pizza with slices of sizes $|p'_i| = |p_i| - x$. If $x > 0$, then for the cutting P' Bob has a strategy with gain strictly greater than $g|P'|/|P|$.*

Proof. Let Σ be Bob's strategy for the cutting P with gain $g \geq |P|/2$. For the cutting P', Bob uses the same strategy Σ. In this way he is guaranteed to get a subset $Q' \subset P'$ of $(n-1)/2$ slices such that the corresponding subset $Q \subset P$ has size at least g. Therefore, Bob's gain is

$$\sum_{p'_i\in Q'} |p'_i| = \sum_{p_i\in Q} \big(|p_i| - x\big) \geq g - x(n-1)/2.$$

Since $|P'| = |P| - xn$, we have to show that $\big(g - x(n-1)/2\big)|P| > g\big(|P| - xn\big)$, which is equivalent to $x(n-1)|P|/2 < xng$. The last inequality follows directly from the assumptions $g \geq |P|/2$ and $x > 0$. ■

Claim 24. *Corollary 20 describes, up to scaling, rotating and flipping the pizza upside-down, all the pizza cuttings into at most 21 slices of at most two different sizes for which Bob has a strategy with gain $5|P|/9$.*

Proof. Let P be a cutting of the pizza into $n \leq 21$ slices of at most two different sizes for which Bob has a strategy with gain $5|P|/9$. If n is even, then Alice has a (zero-jump) strategy with gain $|P|/2$. If all slices have positive size, then by Theorem 1 and Lemma 23, Bob has no strategy with

gain $5|P|/9$. Therefore n is odd and at least one slice in P has size 0. So we can without loss of generality assume that each slice has size 0 or 1.

Now we proceed exactly as in the proof of Claim 22, up to the point where we are showing that $e = 0$. After we apply Claim 21 (three times), we only conclude that $l(B), l(D), l(F) \geq 2$ and consequently $l(A), l(C)$, $l(E) \geq 3$. The length of the circular sequence V' obtained by concatenating arcs B and E is then at most 11. The rest of the argument that $e = 0$ is exactly the same as in the proof of Claim 22. We also conclude that $f = 4|P|/9, d = 3|P|/9$ and $b = 2|P|/9$.

Since the numbers f, d and b are non-negative integers, their sum must be a positive multiple of 9. Since $l(F) + l(D) + l(B) \leq 12$ and the size of each arc is bounded by its length, we have $|P| = 9$, $f = 4$, $d = 3$ and $b = 2$. Consequently, $l(F) \geq 4$, $l(D) \geq 3$, $l(B) \geq 2$ and $l(C) \geq 5$, $l(A) \geq 4$, $l(E) \geq 3$. Since $l(A) + \cdots + l(F) \leq 21$, none of these six inequalities may be strict. Therefore $A = 0000$, $B = 11$, $C = 00000$, $D = 111$, $E = 000$ and $F = 1111$. ■

6. One-Jump Strategies

The main aim of this section is to prove Theorem 3.

The following corollary proves Theorem 3(a).

Corollary 25. *Alice has a zero-jump strategy for V with gain $|V|/3$. The constant $1/3$ is the best possible.*

Proof. The gain $|V|/3$ trivially follows from Observation 11.

Let $V = 100100100$. For every element v of V there is a half-circle C_v of size not greater than $|V|/3$ covering it. So no matter which element v Alice takes in the first turn, as Alice only makes shifts, Bob can play in such a way that Alice gets C_v. ■

In the rest of this section we prove Theorem 3 (b).

6.1. Lower bound

In this subsection we show the strategy for Alice to gain at least 7/16 of the pizza.

We can assume that the number of slices is odd and that $p(V) < |V|/2$ (otherwise Alice has a strategy with gain $|P|/2 \geq 7|P|/16$).

We also fix a minimal triple of half-circles and a partition of V into arcs A, B, C, D, E and F given by Claim 12.

We can use the zero-jump strategy with gain equal to $p(V)$ and the one-jump strategies from Claim 13 (and its analogues). It can be shown, however, that these strategies alone guarantee Alice only 3/7 of the pizza.

To improve Alice's gain we introduce one more one-jump strategy.

Claim 26. *Alice has a one-jump strategy with gain* $3b/8+e/2+\min\{c+d, f+a\}$ *if* $p(V) < |V|/2$.

Proof. If $3b/8 - e/2 \leq 0$, Alice starts by taking a slice from E and then she makes shifts only. As observed in the proof of Claim 13, the potential of any slice in E (and thus Alice's gain) is equal to $e + \min\{c+d, f+a\} \geq 3b/8 + e/2 + \min\{c+d, f+a\}$.

For the rest of the proof we assume that $3b/8 - e/2 > 0$. The main idea of the Alice's strategy is to start with taking a slice somewhere in the arc B and to jump at some appropriate moment before crossing the boundary of B.

Let $k = l(B)$. For $i = 0, 1, \ldots k$, let B_i be the initial subarc of B of length i. Symmetrically, let B'_i be the arc containing the last i slices of B. Similarly we define arcs E_i and E'_i for $i = 0, 1, \ldots k+1$. For $i = 0, 1, \ldots k$, let $h(i) = |B_i| - |E_i|$ and $h'(i) = |B'_i| - |E'_i|$.

The functions h and h' can be used to measure the difference between Alice's and Bob's gain during the first phase (before Alice decides to jump). We call this difference an *advantage* of Alice. During the first phase, Alice takes a sub-arc of B and Bob takes an equally long subarc of E. If Alice took $B_j \setminus B_i$ and Bob took $E_j \setminus E_i$ during the first phase, then Alice's advantage is $h(j) - h(i)$. The other possibility is that Bob took $E_{j+1} \setminus E_{i+1}$; equivalently, Alice took $B'_{i'} \setminus B'_{j'}$ and Bob took $E_{i'} \setminus E_{j'}$, where $i' = k - i$ and $j' = k - j$. In this case the advantage of Alice is $h'(j') - h'(i')$.

By the minimality of the triple that determined the arcs $A, B, \ldots, F$, both functions h and h' are non-negative.

Similarly as in the previous strategies, Bob's best choice after Alice's jump is to let Alice take the rest of the arc E and one of the arcs CD or FA. It follows that if Alice's advantage is g, then her gain will be at least $g + e + \min\{c + d, f + a\}$. It only remains to show that Alice can always achieve an advantage greater than or equal to $3b/8 - e/2$.

Let i be the largest index such that $h(i) \leq 3b/8 - e/2$. Symmetrically, Let i' be the largest index such that $h'(i') \leq 3b/8 - e/2$. We distinguish two cases.

Case 1: $i + i' \leq k$. Equivalently, B_i and $B'_{i'}$ are disjoint. Observe that we actually have $i + i' \leq k - 1$. Alice starts by taking any of the slices from $B \setminus (B_i \cup B_{i'})$. She jumps as soon as Bob takes the first or the last slice from E.

During the first phase either Alice took B_j and Bob took E_j for some $j \geq i + 1$, or Alice took $B'_{j'}$ and Bob took $E'_{j'}$ for some $j' \geq i' + 1$. Alice's advantage is $g = h(j) > 3b/8 - e/2$ in the first case and $g = h'(j') > 3b/8 - e/2$ in the second case.

Case 2: $i + i' > k$. Divide the arc B into consecutive arcs $B^1 = B_{k-i'}$, $B^2 = B_i \setminus B_{k-i'}$ and $B^3 = B'_{k-i}$. Similarly the arc E is divided into $E^1 = E_{k-i'+1}$, $E^2 = E_i \setminus E_{k-i'+1}$ and $E^3 = E'_{k-i+1}$.

Since $|E^2| \geq |E^2| - |B^2| = -\big(h(i) + h'(i') - (b - e)\big) \geq b - e - 2(3b/8 - e/2) = b/4$, we have $\min\big(|E^1|, |E^3|\big) \leq (e - b/4)/2$. We can without loss of generality assume that $|E^1| \leq |E^3|$, hence $|E^1| \leq (e - b/4)/2$. Note that the arc B^1 (and hence E^1) is non-empty, as $|B^1| \geq |B^1| - |E^1| = b - e - \big(|B'_{i'}| - |E'_{i'}|\big) = b - e - h'(i') \geq 5b/8 - e/2 > b/4 > 0$.

Alice now plays as in the proof of Claim 13, where B is replaced by B^1 and E is replaced by E^1. That is, she starts with taking the median slice of B^1 and jumps as soon as Bob takes the first or the last slice of E^1. In this way she gets an advantage

$$g \geq |B^1|/2 - |E^1| = \big(|B^1| - |E^1|\big)/2 - |E^1|/2 = \big(b - e - h'(i')\big)/2 - |E^1|/2 \geq (5b/8 - e/2)/2 - (e - b/4)/4 = 3b/8 - e/2. \quad \blacksquare$$

Remark. By iterating the strategy from Claim 26 we obtain an infinite sequence $\Sigma_1, \Sigma_2, \ldots$ of one-jump strategies, where Σ_{k+1} recursively uses Σ_k in the same way as the strategy Σ_1 from Claim 26 used the strategy Σ_0 from Claim 13. These iterated strategies give better gain when the ratio b/e tends to 1. However, if the ratio b/e is smaller than 5/3, the strategies Σ_i are beaten by the previous one-jump strategies that start outside the arcs B and E.

Proof of the lower bound in Theorem 3(b). As in the proof of Theorem 1, we may without loss of generality assume that $a+b+c \le c+d+e \le e+f+a$.

By Observation 11, Alice has a zero-jump strategy with gain

$$g_1 := e + f + a.$$

By Claim 15, Alice has a one-jump strategy with gain

$$g_2 := b/2 + \min\{c+d, f+a\} = b/2 + c + d.$$

By an analogue of Claim 15, Alice also has a one-jump strategy with gain

$$g_3 := f/2 + \min\{a+b, d+e\} = f/2 + a + b.$$

By Claim 26, Alice has a one-jump strategy with gain

$$g_4 := 3b/8 + e/2 + \min\{c+d, f+a\} = 3b/8 + e/2 + c + d.$$

One of these four strategies gives gain

$$\max\{g_1, g_2, g_3, g_4\} \ge (5g_1 + 3g_2 + 4g_3 + 4g_4)/16$$

$$= (9a + 7b + 7c + 7d + 7e + 7f)/16 = \big(7|P| + 2a\big)/16 \ge 7|P|/16. \qquad \blacksquare$$

6.2. Upper bound

Claim 27. *If Alice is allowed to make only one jump, then Bob has a strategy with gain* $9|P|/16$. *This gain is achieved for the following cutting of the pizza into* 23 *slices:* 20200200202006060050500.

Proof. The characteristic cycle V is depicted on Figure 9. The size of the pizza is 32, which means that we need to show that Bob can get slices with the sum of the sizes at least 18.

The potential of the cutting is 14, thus Alice has no zero-jump strategy with gain greater than 14.

It is easy to check that on the cutting sequence P, every zero slice has a neighboring nonzero slice and both neighbors of every nonzero slice are zero slices. If Alice starts by taking some zero slice, then Bob takes the nonzero

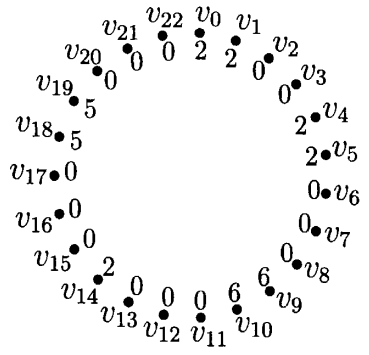


Fig. 9. Characteristic cycle of the cutting used in Claim 27

slice. Alice can then jump, but since she cannot make any more jumps, she would get only at most the potential of V. If she does not jump, she takes a zero slice in her next turn and from Observation 9, Bob can take one of the two arcs. Since Bob already took a nonzero slice, the sum of the sizes of the two arcs is at most 30. If both the arcs have size 15, then both must contain a slice of size 5 because there are no other slices of odd size. But this is impossible because the two slices with size 5 are neighbors on the characteristic cycle. Thus one of the arcs has size at most 14 and if Bob chooses the other one, he gets 18 for the whole game.

Now we may assume that Alice starts by taking a nonzero slice. In his first turn, Bob takes a zero slice such that Alice cannot take a nonzero slice in the next turn. In the first phase, Bob makes shifts. The first phase ends after Alice's jump or if Bob's shift would allow Alice to take a nonzero slice in her next turn.

If Alice jumped in the first phase, she would get at most the slice from her first turn plus the potential of one of its two neighbors from the cutting sequence P. But it is easy to verify that this would mean a gain at most 14 for Alice. If Alice did not start in v_4 or in v_{10} and did not jump, then it is easy to verify that after the first phase, one of the two Bob's zero-jump strategies from Observation 9 guarantees Bob gain 18.

If Alice started in v_{10} and the first phase did not end by Alice's jump, then Bob continues making shifts until either Alice jumps or his shift would allow Alice to take the v_{18} slice in her next turn. This is the second phase. If Alice jumped during the second phase then Bob can make sure that she gets at most $|v_{10}| + |v_{14}| + |v_4| + |v_5| = 12$, see Figure 10 (left)). If she did not jump until the end of the second phase, then from Observation 9,

Bob can make sure that Alice gets at most $|v_{10}| + |v_{14}| + |v_9| = 14$, see Figure 10 (right).

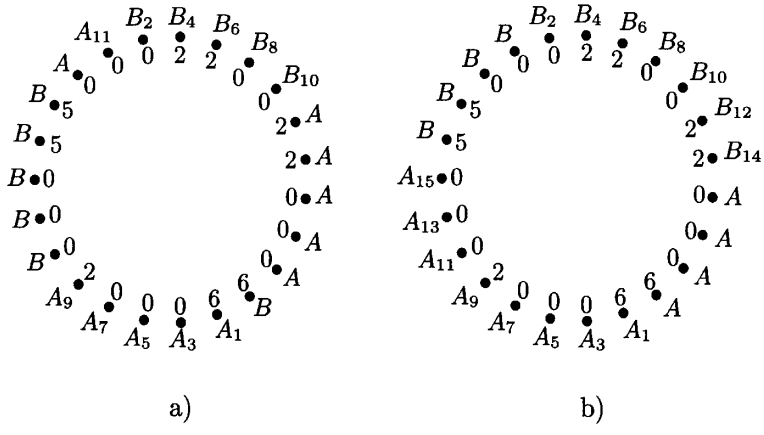


Fig. 10. Two examples of games starting in v_{10} illustrating how Bob can prevent Alice from gaining more than 14

If Alice started in v_4 and the first phase did not end by Alice's jump, then Bob starts the second phase by making a jump and then only shifts. The second phase ends after Alice's jump or if Bob's shift would allow Alice to take v_9. If Alice jumped during the second phase then Bob can make sure that she gets at most $|v_4| + |v_5| + |v_{18}| + |v_{19}| = 14$, see Figure 11 (left). Otherwise we use Observation 9 to show that Bob can make sure that Alice gets at most $|v_4| + |v_5| + |v_0| + |v_1| = 8$, see Figure 11 (right). ■

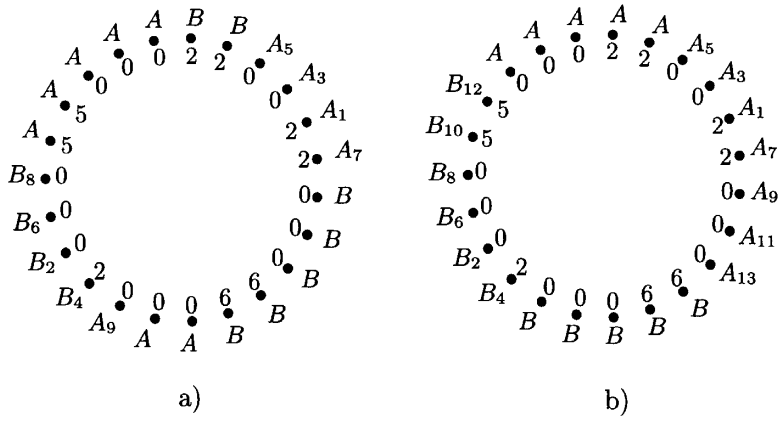


Fig. 11. Two examples of games starting in v_4 illustrating how Bob can prevent Alice from gaining more than 14

7. Algorithms

7.1. Linear Algorithm

In this section we describe an algorithm proving Theorem 6.

For a given cutting of the pizza with n slices, the algorithm computes Alice's two-jump strategy with gain $4|P|/9$ in time $O(n)$.

Without loss of generality we may assume that $|V|$ is a part of the input. The algorithm first computes consecutively the sizes of all n half-circles and finds a half-circle of minimum size in the following way. Consider the characteristic cycle $V = v_0v_1 \dots v_{n-1}$. For $i = 0$ to $n-1$ let s_i be the variable in which the size of the half-circle with starting point v_i, continuing in clockwise direction, is stored. Let s be the size of a currently minimal half-circle H, and v the starting point of H. In the initialization step, compute $s_0 := \sum_{j=0}^{(n-1)/2} |v_j|$ and set $s := s_0$ and $v := v_0$. Then for $i = 1, 2, \dots, n-1$, compute $s_i := s_{i-1} - |v_{i-1}| + \left|v_{(n+2i-1)/2}\right|$. If $s_i < s$, then set $s := s_i$ and $v := v_i$. The above computations can be done in time $O(n)$.

After these precomputations we get a half-circle H of minimum size that we fix. Let v_k and $v_{k+(n-3)/2}$ be the two uncovered elements neighboring H on V. In the following the algorithm computes the potentials of the elements of the uncovered arc $X = v_k \dots v_{k+(n-3)/2}$. Any half-circle covering an element v_i of X also covers v_k or $v_{k+(n-3)/2}$. Let *the right potential* of v_i be the minimum of the sizes of half-circles covering both v_i and $v_{k+(n-3)/2}$. The algorithm computes the right potential p_r for v_k by comparing the values of s_{k-1} and s_k, i.e., $p_r(v_k) := \min\{s_{k-1}, s_k\}$. For $i = k+1$ to $k+(n-3)/2$ set $p_r(v_i) := \min\left\{p_r(v_{i-1}), s_i\right\}$. Analogously let *the left potential* of v_i be the minimum of the sizes of half-circles covering both v_i and v_k. The computation of the left potentials p_l is similar. Obviously the potential of v_i is $\min\left\{p_l(v_i), p_r(v_i)\right\}$. The computations are done in time $O(n)$.

The potential of any element of X is at least as big as the potential of any element of H due to the minimality of H. Therefore $p(V)$ is equal to the maximal potential on X. If $p(V) \geq |V|/2$, then the algorithm returns an element of potential $p(V)$ in time $O(n)$. This will be Alice's first turn and all her other turns will be shifts that can be computed in time $O(1)$.

From now on we assume that $p(V) < |V|/2$. The algorithm finds the index $j \in X$ for which $p_l(v_j) + p_r(v_j)$ is minimal among all j such that

both $p_l(v_j)$ and $p_r(v_j)$ do not exceed $p(V)$. There exists such a $j \in X$ as a consequence of Claim 10 and Claim 12. Let H_1 be the half-circle that gives the left potential $p_l(v_j)$ for v_j, and H_2 the half-circle that gives the right potential $p_r(v_j)$ for v_j. Then H, H_1 and H_2 form a minimal triple. Indeed, suppose that there is a half-circle in the triple that can be replaced by a half-circle of strictly smaller size. Clearly, this half-circle is not H but H_1 or H_2. A contradiction to the choice of j. We get that the triple is minimal.

Knowing the minimal triple the algorithm computes A, B, C, D, E, F and V' in time $O(n)$. If $p(V') \geq |V'|/2$, a slice of potential $p(V')$ can be found in time $O(n)$. Otherwise, the algorithm computes the arcs A', B', C', D', E', F' on V' similarly as above in time $O(n)$. A median slice of B' can be found in time $O(n)$ by traversing B' twice. At first the algorithm computes $|B'|$. Then it adds the sizes of the elements one by one again and checks in every step if the sum exceeds $|B'|/2$. That will occur at a median slice.

The algorithm orders $a+b+c, c+d+e$ and $e+f+a$. Assume without loss of generality that $a+b+c \leq c+d+e \leq e+f+a$. According to Section 2.5 Alice has three strategies with gains $g_1 := e+f+a, g_2 := b/2+e/4+c+d$ and $g_3 := f/2+c/4+a+b$. These strategies were computed in time $O(n)$ as described above. Alice chooses one of the three strategies corresponding to $\max\{g_1, g_2, g_3\}$. Once the strategy is known, Alice's turn can be computed in time $O(1)$ in any position of the game.

7.2. Optimal strategies

The following result implies, for example, that Alice can be forced to make only jumps (except A_1, A_n) in her optimal strategy.

Observation 28. (i) *For any odd $n \geq 3$ and for any of the $2^{n-2}n$ permutations allowed in the game on n slices, the pizza can be cut into n slices in such a way that if both Alice and Bob make only optimal turns then the order of taken slices is the chosen permutation.*

(ii) *For any even $n \geq 4$ and for any of the $2^{n-3}n$ plausible permutations in which Alice's second turn is a jump, the pizza can be cut into n slices in such a way that if both Alice and Bob make only optimal turns then the order of taken slices is the chosen permutation.*

Proof. We give the sizes $1, 1/2, 1/4, \ldots$ in the order in which we want the slices to be taken. ∎

Now we describe an algorithm that computes both player's optimal strategy for a given cutting of the pizza with n slices in time $O(n^2)$.

Proof of Claim 7. A *position* of a game is an arc X characterized by its leftmost slice x_l and its rightmost slice x_r or the empty-set if there are no more slices left. If $l(X) = 1$, then $x_l = x_r$. There are $n^2 - n + 2$ possible positions X. The parity of $l(X)$ determines whose turn it will be. For $i = 0$ to $n - 1$ the algorithm traverses all X with $l(X) = i$ and decides the best strategy for the player on turn. The best possible gain on X is the *value* of X, denoted by $v(X)$. The algorithm stores $v(X)$ for all positions X. For $i \in \{0, 1\}$ the strategy is obvious and $v(X) = |X|$. Let $X - x$ be the arc X omitting the slice x. For $i \geq 2$, $v(X) = |X| - \min\{v(X - x_j), v(X - x_r)\}$ and the player takes the corresponding slice yielding the minimum in the previous expression. All this can be done in time $O(n^2)$.

A program implementing the algorithm described above can be downloaded from the following webpage: http://kam.mff.cuni.cz/~cibulka/pizza. The program can be used to verify Claims 17, 20 and 27. ■

Acknowledgments. We thank Peter Winkler for posing this problem and for the discussions. The third author thanks Dirk Oliver Theis for the discussions about the upper bound during Fete of Combinatorics and Computer Science in Keszthely, August 11–16, 2008. We also acknowledge the use of the LPSolve IDE [2] software, which made our computations much more efficient.

References

[1] J. Cibulka, J. Kynčl, V. Mészáros, R. Stolař and P. Valtr, *Generalizations of Peter Winkler's Pizza Problem,* in preparation.

[2] H. Gourvest, LPSolve IDE v5.5, http://www.progdigy.com
M. Berkelaar, J. Dirks, K. Eikland and P. Notebaert, lp_solve 5.5, http://lpsolve.sourceforge.net/5.5

[3] K. Knauer, P. Micek and T. Ueckerdt, *How to eat* 4/9 *of a pizza,* manuscript.

Josef Cibulka, Rudolf Stolař

Department of Applied Mathematics,
Charles University,
Faculty of Mathematics and Physics,
Malostranské nám. 25,
118 00 Praha 1,
Czech Republic

e-mail: cibulka@kam.mff.cuni.cz
ruda@kam.mff.cuni.cz

Jan Kynčl, Viola Mészáros, Pavel Valtr

Department of Applied Mathematics and Institute for Theoretical Computer Science,
Charles University,
Faculty of Mathematics and Physics,
Malostranské nám. 25,
118 00 Praha 1,
Czech Republic

e-mail: kyncl@kam.mff.cuni.cz

Viola Mészáros

Bolyai Institute,
University of Szeged,
Aradi vértanúk tere 1,
6720 Szeged,
Hungary

e-mail: viola@math.u-szeged.hu

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 95–137.

TIGHT BOUNDS FOR EMBEDDING BOUNDED DEGREE TREES

BÉLA CSABA*, IAN LEVITT, JUDIT NAGY-GYÖRGY
and ENDRE SZEMERÉDI†

Let T be a tree on n vertices with constant maximum degree K. Let G be a graph on n vertices having minimum degree $\delta(G) \geq n/2 + c_K \log n$, where c_K is a constant. If n is sufficiently large then $T \subset G$. We also show that the bound on the minimum degree of G is tight.

1. INTRODUCTION

In this paper we will only deal with simple graphs, without loops or multiple edges. Given a graph $F = (V, E)$, and $U \subset V$, the subgraph of F induced by the vertices of U will be denoted by $F|_U$. The degree of a vertex $v \in V(F)$ is denoted by $\deg_F(v)$, or by $\deg(v)$ if F is clear from the context. The minimum degree and maximum degree of F will be denoted by $\delta(F)$ and $\Delta(F)$, respectively. We will let $N(v)$ represent the neighborhood of a vertex v, hence $\deg(v) = \left|N(v)\right|$. The number of vertices in a graph F is written $v(F)$, and the number of its edges is $e(F)$. If $A, B \subset V(F)$, then $e(A, B)$ is the number of edges of F with one endpoint in A and the other endpoint in B, and $e(A) = e(A, A) = e\big(F|_A\big)$. For a tree T we denote the set of its leaves by $\ell(T)$.

Given a bipartite graph H, a proper 2-coloring of H is a 2-coloration of the vertices of H such that adjacent vertices of H get different colors.

*Partially supported by the Office of Sponsored Programs at WKU and the OTKA grants T049398 and K76099.

†Partially supported by the Ellentuck Fund.

For graphs J and F, we will write $J \subset F$ if and only if there exists an adjacency preserving injection $I : V(J) \to V(F)$. If $I(x) = v$ then we say that x *maps onto* v and that v *is covered by* x. If a vertex of F is not covered by any vertex of J then we call it *uncovered.*

Throughout the paper we use $a \ll b$ to denote that a is sufficiently smaller than b. Our goal is to show the following:

Theorem 1. *Assume that* $T = \big(V, E(T)\big)$ *is a bounded degree tree on* n *vertices with* $\Delta(T) = K$. *Let* $G = \big(W, E(G)\big)$ *be a graph on* n *vertices. Assume that its minimum degree,* $\delta(G) \geq n/2 + c_K \log n$ *where* c_K *is a constant depending only on* K. *Then there exists* n_0 *such that* $T \subset G$ *for* $n \geq n_0$. *Furthermore, the bound on* $\delta(G)$ *is tight: there exists a graph* G *with* $\delta(G) > n/2 + \log n/17$ *such that the complete ternary tree on* n *vertices is not a subgraph of* G *if* n *is sufficiently large.*

Bollobás [2] conjectured that if G is a graph on n vertices, $\delta(G) \geq (1/2+\varepsilon)n$ for some $\varepsilon > 0$, and T is a bounded degree tree on n vertices, then $T \subset G$. The problem was solved in the affirmative by Komlós, Sárközy and Szemerédi [3] for large graphs. They then strengthened their result (see [4]), and showed, that $\Delta(T)$ need not be bounded: there exists a constant c such that $T \subset G$ if $\Delta(T) \leq cn/\log n$, $\delta(G) \geq (1/2 + \varepsilon)n$ and n is large. Both proofs are based on the Regularity Lemma – Blow-up Lemma Method. In this paper we eliminate the use of the Regularity Lemma under the original assumption that $\Delta(T)$ is bounded. As a result the threshold n_0 is much smaller, but more importantly, we prove a stronger statement than the original conjecture. Indeed, in Theorem 1 $\delta(G) - n/2$ can be as small as $O(\log n)$, and $o(\log n)$ is not sufficient.

We may assume that G is minimal with respect to edge deletion, retaining the property that $\delta(G) \geq n/2 + c_K \log n$. In particular, there are no edges $xy \in E(G)$ such that *both* vertices x, y have degree larger than $n/2 + c_K \log n$.

We divide the problem into two subproblems, depending on whether G is *extremal* or *non-extremal.* We call G γ-extremal for $\gamma > 0$ if either G or its complement $\overline{G}$ contains a subgraph on $n/2$ vertices with at most γn^2 induced edges. Otherwise G is γ-non-extremal. Let $\gamma = K^{-20}$. We first prove the statement for γ-non-extremal graphs, then for γ-extremal graphs.

In each case we will construct the adjacency preserving mapping $I : V \to W$ step by step. It is worth noting that while at any time in the course of embedding we have an adjacency preserving partial mapping, the

final I is not necessarily an extension of it. It is possible that we rethink the way we mapped certain subtrees.

In the last section we show that for complete ternary trees the $\log n$ additive term in the minimim degree of G is necessary in case G is a suitably defined extremal graph. Let us remark that studying some cases of trees having larger degrees suggest that one needs the minimum degree $\delta(G) \geq n/2 + C\Delta(T)\log n$, up to the point when the tree has maximum degree $cn/\log n$ for some positive constant c.

Throughout the paper we make no attempt to optimize on the absolute constants.

2. The Non-Extremal Case

The case of non-extremal G is divided into two subcases depending on the structure of T. It is shown below that T has either a small subtree with many leaves, which we call a *broad subtree*, or it has a large subtree with few leaves, called a *long subtree*. Specifically, T_b is a broad subtree of T if $\ell(T_b) \subset \ell(T)$, $v(T_b) \leq \gamma^5 n/K^2$ and $\left|\ell(T_b)\right| \geq \gamma^7 n$. We call $T_l \subset T$ long if $\ell(T_l) \subset \ell(T)$, $v(T_l) > \gamma^5 n/K^3$ and $\left|\ell(T_l)\right| < \gamma^7 n$. We use a simple method to decide which case applies.

2.1. Some tools for the proofs in the non-extremal cases

Definition 1. Assume that we are given a rooted tree F with root ρ. Let x and y be any two vertices of F. We say that y is *below* x if the simple path connecting y with ρ goes through x. Let $F(x)$ denote the subtree rooted at x containing every vertex which is below x. Sometimes we will call x the tip of the subtree $F(x)$.

We turn to the method of determining whether T has a broad or a long subtree. First, (temporarily) designate a root $\rho \in V(T)$. Repeat the following for every $x \in V(T)$: color x blue if $\left|\ell\big(T(x)\big)\right| \geq \gamma^7 n$, otherwise color it red. If ρ is red, then T itself is long. If not, then T must have a blue vertex x such that all of its children are red. If $v\big(T(x)\big) \leq \gamma^5 n/K^2$ then $T(x)$ is a broad subtree and we let $T_b = T(x)$. If $v\big(T(x)\big) > \gamma^5 n/K^2$

then one of its red children, y, is such that $v\big(T(y)\big) > \gamma^5 n/K^3$. $T(y)$ is a long subtree, and we let $T_l = T(y)$.

Observe that if G is not γ-extremal, then it has many triangles: there are at least γn^2 edges in the neighborhood of every vertex, since these neighborhoods are of size larger than $n/2$. As any such edge is in a triangle, summing over vertices will count each triangle three times. We find that there are at least $\gamma n^3/3$ triangles in G.

We record here a simple statement, which is also very useful throughout the proof:

Lemma 2. *Let H be a tree of size m. If J is a graph with $\delta(J) \geq m$, then there is an embedding of H into J. If $J' = J'(A, B)$ is a bipartite graph such that every $b \in B$ has at least $m/2$ neighbors in A and every $a \in A$ has at least m neighbors in B then $H \subset J'$.*

Proof. We leave the proof of the first part to the reader. For the second part we remark that H is 2-colorable, and one of its color classes has size at most $m/2$. ■

We will frequently apply the following folklore statement (proof is omitted):

Lemma 3. *Every graph H has a subgraph H' such that*

$$\delta(H') \geq e(H)/v(H). \qquad \blacksquare$$

Remark 1. We will frequently use the above two lemmas in the following way: whenever we find a dense subgraph in the uncovered part of W, we can find a subgraph of it with large minimum degree into which we can map a large subtree of T.

Let $F = F(A, B)$ be a bipartite graph satisfying the following requirements:

- $|A| = t$ and $|B| = t^2$,
- $\gamma^{-3} \ll t$,
- every $b \in B$ has at least $(1/2 + \gamma^3)t$ neighbors in A.

Then we have the following *Cleaning Lemma.*

Lemma 4 (Cleaning lemma – first version). *F has a subgraph $F' = F'(A', B')$ such that $A' \subset A$, $B' \subset B$, every $b \in B'$ has at least $(1/2+\gamma^3/2)t$ neighbors in A' and every $a \in A'$ has at least t neighbors in B'.*

Proof. First we assume, that every vertex in B has exactly $(1/2 + \gamma^3)t$ neighbors in A – if necessary, we discard edges incident to those vertices of B which have larger degrees. Then we have $e(F) = (1/2 + \gamma^3)t^3$. We will find the desired subgraph F' step-by-step in the following way.

Let $A_1 = \{a \in A : |N(a) \cap B| < t\}$, and let $B_1 = \{b \in B : |N(b) \cap (A - A_1)| \leq (1/2 + \gamma^3/2)t\}$. In the first *cleaning step* we delete the vertices of A_1 from A and the vertices of B_1 from B. Removing these vertices deletes at most $|A_1|t + 2|A_1|t(1/2 + \gamma^3)/\gamma^3$ edges from F.

In the ith cleaning step we first identify a subset $A_i \subset A - \bigcup_{j=1}^{i-1} A_j$ and a subset $B_i \subset B - \bigcup_{j=1}^{i-1} B_j$:

$$A_i = \left\{ a \in A - \bigcup_{j=1}^{i-1} A_j \, : \, \left| N(a) \cap \left(B - \bigcup_{j=1}^{i-1} B_j \right) \right| < t \right\}$$

and

$$B_i = \left\{ b \in B - \bigcup_{j=1}^{i-1} B_j \, : \, \left| N(b) \cap \left(A - \bigcup_{j=1}^{i} A_j \right) \right| < (1 + \gamma^3)t/2 \right\}.$$

Clearly $e\big(A_i, B - \bigcup_{j=1}^{i-1} B_j\big) < |A_i|t$, and moreover,

$$|B_i| \leq 2\frac{e\big(A_i, B - \bigcup_{j=1}^{i-1} B_j\big)}{\gamma^3 t} < 2\frac{|A_i|}{\gamma^3}.$$

Deleting the vertices of A_i and B_i therefore removes at most $|A_i|t + 2|A_i|(1/2 + \gamma^3)t/\gamma^3$ edges. Let us assume that after k cleaning steps the cleaning process stops: either every vertex left satisfies the degree requirements of the lemma, or there are no vertices left. The total number of edges we lose from F is at most

$$\sum_{1 \leq i \leq k} |A_i|\big(t + 2t(1/2 + \gamma^3)/\gamma^3\big) < t^2 \left(1 + \frac{1 + 2\gamma^3}{\gamma^3}\right) = t^2(3 + 1/\gamma^3) \ll t^3.$$

We must have the former case, as most of the edges are still present when the process stops. The induced subgraph on $A - \bigcup_1^k A_i$ and $B - \bigcup_1^k B_i$ will be denoted by F', and is easily seen to satisfy the requirements of the lemma. ∎

For the second version of the Cleaning Lemma we assume that F satisfies the following:

- $|A| = t$ and $|B| = t^3$,
- $\gamma^{-4} \ll t$,
- every $b \in B$ has at least $(1/2 - \gamma^4)t$ neighbors in A.

While the setup is somewhat different, the proof of the Lemma is very similar to that of the first version, we omit the details:

Lemma 5 (Cleaning lemma – second version). *F has a subgraph $F' = F'(A', B')$ such that $A' \subset A$, $B' \subset B$, every $b \in B'$ has at least $(1/2 - 2\gamma^4)t$ neighbors in A' and every $a \in A'$ has at least t^2 neighbors in B'.* ∎

Randomly choose three disjoint subsets from W, denoted by D_1, D_2 and M, with sizes $|D_1| = |D_2| = \gamma^8 n$, and $|M| = \gamma^{10} n$. The common size of the D_is will be denoted by s. We make three assertions regarding these sets:

- Every $u \in W$ has at least $s/2 - O\big(\sqrt{n \log n}\big)$ neighbors in the D_is and $|M|/2 - O\big(\sqrt{n \log n}\big)$ neighbors in M with high probability.
- If G is γ-non-extremal, then the subgraphs induced by D_1, D_2 and M will be $\gamma/2$-non-extremal with high probability.
- The induced bipartite subgraph on D_1 and D_2 will be $\gamma/2$-non-extremal with high probability. That is, if $A \subset D_1$, $B \subset D_2$, each of size $s/2$, then there are at least $\gamma s^2/2$ edges between A and B.

The proofs of these statements are straightforward. We sketch a proof of the non-extremality of D_i. We demonstrate that the induced subgraph on D_1 cannot have a sparse subset of size $s/2$. The analogous statement, that the complement of the above induced subgraph cannot have a large subset with a few edges can be proved similarly.

It is a simple exercise that if G is not extremal, then any set U for which $\forall a, b \in U$, $\big|N(a) \cup N(b)\big| \leq (1/2 + \gamma)n$ cannot contain more than $(1/2 - \gamma)n$ vertices. Conversely, if G were extremal, it would contain a set U of size $n/2$ and an element $a \in U$ such that $\big|N(a) \cup N(b)\big| \leq (1/2 + \gamma)n$ for all but at most γn vertices $b \in U$. Let $A \subset D_1$ be any set of size $s/2$. If $\exists a \in A$ such that $\big|N(a) \cup N(b)\big| \leq (1/2 + \gamma)n$ for most $b \in A$ then with high probability about half the vertices of $W(G)$ would have almost the same neighborhood

as a has. But that would imply the extremality of G. Thus, with high probability, a set $A \subset D_1$ of size $s/2$ will have the property that for many pairs of vertices $a, b \in A$,

$$\left|\left(N(a) \cup N(b)\right) \cap D_1\right| \geq (1/2 + \gamma)s - O\left(\sqrt{n \log n}\right).$$

It follows that for most pairs of vertices, $\left|\left(N(a) \cup N(b)\right) \cap A\right| \geq \gamma s/2$. Thus A is dense.

Lemma 6. *Let G be a non-extremal graph, and D_1, D_2 be as above. Then there is a perfect matching in the bipartite subgraph of G induced by D_1 and D_2.*

Proof. We will check the König–Hall conditions. Since every $u \in W$ has at least $s/2 - o(n)$ neighbors in D_2, every $H \subset D_1$ such that $|H| \leq s/2 - \gamma^2 s$ will have a neighborhood of size at least $|H|$ in D_2. Assume that $|H| > s/2 - \gamma^2 s$. If D_2 had a subset $\widehat{H}$ of size $s/2 - \gamma^2 s$ such that $e(H, \widehat{H}) = 0$, then the induced bipartite subgraph on D_1 and D_2 would be extremal – if we complete H and $\widehat{H}$ by adding at most $\gamma^2 s$ new vertices to both, the number of newly added edges is not more than $2\gamma^2 s^2 \ll \gamma s^2/2$. Hence every $H \subset D_1$ having size at least $s/2 - \gamma^2 s$ has a neighborhood of size at least $s/2 + \gamma^2 s$ in D_2. We finish by noticing that every vertex in D_2 has $s/2 - o(n)$ neighbors, therefore, if $H \subset D_1$ is larger than $s/2 + o(n)$ then it is neighboring with the whole D_2. ∎

The following corollary is used to handle a small modification of the bipartite subgraph from Lemma 6.

Corollary 7. *Discard at most $\gamma^2 s$ vertices from D_2, replacing them with arbitrary vertices which all have at least $|D_1|/2 - o(n)$ neighbors in D_1. Then the resulting new graph has a perfect matching.*

Proof. The proof of Lemma 6 is based on two facts: the first is that every vertex in the bipartite graph has at least $s/2 - o(s)$ neighbors in the other color class, the second is that the bipartite subgraph induced by D_1 and D_2 is non-extremal. Since every vertex of G has at least $s/2 - o(s)$ neighbors in D_1, if we replace $\gamma^2 s$ vertices of D_2 by any $\gamma^2 s$ vertices, the first condition will remain satisfied. For the second observe, that the modified bipartite graph will be $\gamma/3$-non-extremal, since we lost at most $\gamma^2 s^2$ edges between any two subsets of size $s/2$ when discarding $\gamma^2 s$ vertices from D_2. ∎

Lemma 8. *Assume that G is non-extremal and $u, v \in W$. Then there are at least $\gamma n/5$ vertex disjoint paths of length 3 connecting u and v.*

Proof. Set $a = \big|N(u) \cap N(v)\big|$. If $a \geq n/2 - \gamma n/2$ then, by the non-extremality of G, there are at least $\gamma n^2/2$ edges in $N(u) \cap N(v)$. One can therefore find $\gamma n/2$ vertex disjoint edges in $N(u) \cap N(v)$, determining $\gamma n/2$ vertex disjoint paths of length 3 between u and v. At the other extreme, if $a \leq \gamma n/2$ then there are at least γn^2 edges in $N(u)$ induced by $\overline{G}$, again since G is non-extremal. If v' is the endpoint of an edge in $\overline{G}|_{N(u)}$, then it has a neighbor in $N(v)$. It follows that there are at least $\gamma n^2/2$ edges between $N(u)$ and $N(v)$, and we can easily find the vertex disjoint paths of length 3 between u and v.

Assume now that $\gamma n/2 \leq a \leq n/2 - \gamma n/2$. Then the number of edges connecting $N(u) \cap N(v)$ with $N(u) \cup N(v)$ is at least $a(n/2 - a) \geq \gamma n^2/5$. From these one can easily choose $\gamma n/5$ vertex disjoint edges to yield the desired vertex disjoint paths of length 3. ∎

2.2. The first non-extremal case: T has a broad subtree

2.2.1. Preparations

Finding distinguished leaves. Recall that $T_b \subset T$ is a broad subtree with $\ell(T_b) \subset \ell(T)$, $\big|\ell(T_b)\big| \geq \gamma^7 n$ and $v(T_b) \leq \gamma^5 n/K^2$. We choose the tip of T_b to be the root of T, and denote it by r. As T has bounded degree, we can easily find $s = \gamma^8 n$ subtrees of T_b, denoted by $\{F_1, F_2, \ldots, F_s\}$, such that each F_i has depth 4, and for each i, j, the distance between F_i and F_j is at least 4. Furthermore, for each $1 \leq i \leq s$ we distinguish a leaf $a_i \in \ell(F_i) \cap \ell(T_b)$ which is at depth 4. The immediate ancestor of a_i is denoted by $P(a_i)$ and the root of F_i is called $A(a_i)$. Observe that $P(a_i)$ is distance three from $A(a_i)$. Denote the set of distinguished leaves by $S = \{a_1, a_2, \ldots, a_s\}$. We will map $P(a_1), P(a_2), \ldots, P(a_s), a_1, a_2, \ldots, a_s$ in a special way. In particular, we will map the $P(a_i)$s to the vertices of D_1, and most of the a_is to D_2.

Embedding T_b. We start the embedding of T by that of T_b. We first map the $A(a_i)$s into $W - (D_1 \cup D_2 \cup M)$ arbitrarily, and the $P(a_i)$s into D_1 arbitrarily. Using Lemma 8 we find vertex disjoint paths of length 3 connecting $I\big(A(a_i)\big)$ with $I\big(P(a_i)\big)$, such that these paths avoid D_2 and M, and contain vertices from D_1 only as an endpoint. We then finish the embedding of T_b, avoiding $M \cup D_2$. Since T_b is very small and the minimum

degree is large in $G \setminus (M \cup D_2)$, we can easily embed in a bottom-up fashion from the $A(a_i)$s towards r and top-down from the $A(a_i)$s towards the leaves. We delay the embedding of S until the end. Observe, that with the above we have determined I on $V(T_b) - S$.

2.2.2. Decomposition of $T - T_b$. Our goal is to decompose $T - T_b$ into subtrees of constant size such that most of the vertices are in *good* subtrees. We call a subtree in the decomposition good if there are at most two edges which connect it to other subtrees. We make the following claims, whose proofs are left to the reader:

Claim 9. *Let F be a tree. The number of vertices of F having degree at most two is at least $v(F)/2$.*

Claim 10. *Let F be a tree. There are at most $\frac{v(F)}{\alpha-2}$ vertices which have degree at least α.*

The basic building block of the decomposition algorithm is simple. Observe that given a rooted tree F with maximum degree K and an integer $h \leq v(F)$ one can find a vertex $x \in F$ such that $h \leq v\big(F(x)\big) < Kh$. Decompose F by trimming off subtrees recursively as long as there are at least $(K+1)h$ vertices not yet part of the decomposition. When the process stops, the remaining connected component comprises the final subtree in the decomposition. We call such a decomposition of F into subtrees all having size in the range $\big[h, (K+1)h\big]$ an *h-decomposition.* Construct an auxiliary tree $\widetilde{F}$ whose vertex set is the set of the subtrees in the decomposition of F, and for which two vertices are connected iff there is an edge between the corresponding subtrees. By applying Claim 9 to $\widetilde{F}$, we conclude that at least $v(F)/\big(2(K+1)\big)$ vertices of F are in good subtrees.

We want to discard those subtrees of the decomposition which have many neighbors in $\widetilde{F}$. Call a subtree *bad* if the corresponding vertex has degree larger than $2K\gamma^{-10}$ in $\widetilde{F}$. By Claim 10 F has at most $v(\widetilde{F})/(2K\gamma^{-10}-2)$ bad subtrees. It follows that the total number of vertices of F in bad subtrees is at most $\gamma^{10}v(F)/2$.

Set $m = 3(K+1)\log\gamma^{-10}$ and let $h_1 > h_2 > \ldots > h_m$ be a sequence of integer constants, such that $h_1 = \gamma^{-40}\big(4K^3\gamma^{-10}\big)^m$, and $h_i/h_{i+1} = 4K^3\gamma^{-10}$. For the full decomposition of $T - T_b$, we iteratively determine h_i-decompositions of smaller and smaller subtrees. In the process, we define a sequence of connecting edge sets, the first of which is $E_1 = \emptyset$. First, find

an h_1-decomposition of $T - T_b$ and take the union of $T - T_b$ with E_1. The set of subtrees in the h_1-decomposition is divided into three sets – the good, the bad, and the active. Good and bad subtrees are as defined above, and the remaining subtrees are called *active*. Let E_2 be the edges connecting the subtrees of the decomposition.

For each active subtree F we find an h_2-decomposition of F. We determine the h_2-good subtrees of this decomposition as those which are connected to at most 2 other subtrees of F, noting that they may also be connected through edges of E_2 to the larger decomposition. The bad subtrees determined by the decomposition of F are those which are connected to at least $2K\gamma^{-10}$ other subtrees, either within F or through E_2. The rest of the subtrees are considered active for the next iteration. Having found an h_2-decomposition of all of the active subtrees, define E_3 to be the edges connecting all of the subtrees thus far determined – i.e. E_3 is the union of E_2 and the edges connecting the subtrees of the all of the h_2-decompositions of the active subtrees from the previous step.

In general, after finding the h_i-decomposition of every active subtree from the h_{i-1}-decompositions and distinguishing the h_i-good, the bad, and the active subtrees as before, we define E_i to be the union of E_{i-1} and the set of edges connecting the subtrees of the h_i-decompositions. We stop after the m^{th} iteration.

We are interested in subtrees that are good in the entire decomposition. We claim that the number of vertices in such good subtrees is at least $(1 - \gamma^{10})v(T - T_b)$. Indeed in the i^{th} step, having decomposed an active subtree into N subtrees, at least $N/2$ of them are h_i-good. Since E_i has at most $2K\gamma^{-10}$ edges which are adjacent to the decomposed active subtree, at most $2K\gamma^{-10}$ of the h_i-good subtrees out of $N/2$ are not good overall. By the definition of the h_i at least $N/3$ subtrees will remain good, and these will contain at least $h_i/3(K+1)$ vertices of the decomposed active subtree. That is, at every step we gain a proportion of $1/\big(3(K+1)\big)$ for the new vertices in good subtrees.

Since $\big(1 - 1/\big(3(K+1)\big)\big)^m \leq \gamma^{10}/2$, in the final decomposition at most $\gamma^{10}v(T - T_b)$ vertices are not in good subtrees, and every subtree has size between γ^{-40} and $\gamma^{-40}\big(4K^3\gamma^{-10}\big)^m$.

Certain good subtrees won't be easy to handle later. Consider a good subtree $\hat{T}$ which is connected to two other subtrees in the decomposition such that these two connecting edges have a common endpoint x in $\hat{T}$. In order to eliminate this problem we do the following. First, delete $\hat{T}$

from the set of good subtrees. Let $y_1, y_2, \ldots, y_s$ denote the neighbors of x in $\hat{T}$ and let $\hat{T}(y_1), \hat{T}(y_2), \ldots, \hat{T}(y_s)$ denote the subtrees of $\hat{T}$ originating from $y_1, y_2, \ldots, y_s$, respectively. If $v\big(\hat{T}(y_i)\big) > \gamma^{-20}$ then $\hat{T}(y_i)$ will be a good subtree, otherwise it will be bad. Since $v(\hat{T}) \geq \gamma^{-40}$ and at most $(K-1)\gamma^{-20}$ vertices are put into the set of bad subtrees, we still have at least $(1-\gamma^{10})v(T-T_b)$ vertices in good subtrees.

Denote the good subtrees of $T - T_b$ by $\hat{T}_1, \hat{T}_2, \ldots, \hat{T}_\ell$. Each good subtree $\hat{T}_i$ has bounded size, there are at most two edges which connect the good subtrees to any other subtree in the decomposition, the endpoints of these connecting edges always differ, and the vast majority of the vertices are in the union of the good subtrees.

2.2.3. Sketch of the embedding algorithm. We map large subtrees of $T - T_b$ one-by-one using the Main Mapping Procedure, discussed below. This procedure works smoothly until about half of $T - T_b$ is mapped at which point we are faced with problems satisfying the key constraint of the Main Mapping Procedure, namely that the uncovered vertices of G contain a large subgraph with a non-negligible minimum degree. If the set of uncovered vertices is too sparse, then we rethink the way we mapped certain subtrees. We will *rearrange* the mapping of some subtrees in such a way that we use many uncovered vertices, and the new set of uncovered vertices contains a large dense bipartite graph, which we use to continue the embedding of T.

When at most $\gamma^{10}n$ vertices are left out from $T - T_b$, we can easily finish the embedding.

2.2.4. Finishing the embedding of T. Let us assume that we have already mapped most of T, only the vertices of S and a few subtrees, $T_1, T_2, \ldots, T_k$ are left unmapped, such that $|\cup T_j| \leq \gamma^2 s$. Set $T^* = T - (\cup T_j \cup S)$. We embed the T_j into D_2, easily accomplished as D_2 is a randomly chosen subset of W, and $|D_2| = s \gg \gamma^2 s$. A perfect matching between D_1 and $D_2 - I(\cup T_j) \cup \big(W - \big(I(T^*) \cup D_2\big)\big)$ is ensured by Corollary 7 and determines the rest of the embedding. Indeed, as D_1 is covered by the $P(a_i)$, we map a_i to the vertex matched to $I\big(P(a_i)\big)$. Hence, if we can map most of T such that we avoid D_2, we can finish the embedding of T. In the leftover of this section we will work so as to achieve this goal.

2.2.5. The Main Mapping Procedure. The method below is used to map large chunks of T into G one-by-one. We apply it throughout the embedding.

Assume that we are in the course of the embedding. Suppose that the leftover of T has a subtree $\widetilde{T}$ of the following type: its root is the already mapped vertex y, y's children are $y_1, y_2, \ldots, y_m$ ($m \leq K$), the children of the y_is is the set $\{r_1, r_2, \ldots, r_\ell\}$, and there is a tree T_j of constant size originating at r_j for every $1 \leq j \leq \ell$. Denote by H a subset of those vertices of $W - (M \cup D_1 \cup D_2)$ which have not yet been covered by a vertex of T. Assume that every vertex of H has at least $\gamma^{-2}v(\widetilde{T})$ neighbors in H. Then we claim the following:

Lemma 11. *Assume that the above conditions are satisfied. If M has at least $(1-\gamma/10)|M|$ vacant vertices, then I can be extended to the vertices of $\widetilde{T}$ such that $I(y_i) \in M$, $I(T_j - r_j) \subset H$, and $I(r_j) \in M$ for every $1 \leq i \leq m$ and $1 \leq j \leq \ell$.*

Proof. Let us denote the set of already covered vertices of M by X, then $|X| \leq \gamma|M|/10$. Since M is a randomly chosen set, $\big|N\big(I(y)\big) \cap (M - X)\big| \geq (1/2 - \gamma/9)|M|$. Let us define the set

$$U = \big\{u : u \in M - X,\ \big|N(u) \cap H\big| \geq \gamma^2|H|\big\}.$$

Again, as M is randomly chosen, every vertex of H has at least $(1/2 - \gamma/9)|M|$ neighbors in $M - X$. Simple computation shows that $|U| \geq (1/2 - \gamma/8)|M|$. Since the induced subgraph on M is $\gamma/2$-non-extremal, there are at least $\gamma|M|^2/5$ edges in between U and $N\big(I(y)\big) \cap (M - X)$. Then we map the y_is to such vertices in $N\big(I(y)\big) \cap (M - X)$ which have at least K^2 neighbors in U. If $y_i r_k$ is an edge in T, then r_k will be chosen among the neighbors of $I(y_i)$. It is possible to embed the remaining small T_j subtrees in H due to the assumed minimum degree of H. ∎

Remark 2. We remark that when embedding such a subtree, we cover less than K^2 vertices from M. Therefore, if one embeds at least γ^{-12} vertices in each application of the Main Mapping Procedure, less than $\gamma|M|/10$ vertices of M will be covered throughout the embedding.

Remark 3. We will apply the Main Mapping Procedure throughout the embedding. Whenever the set of vacant vertices contains a large dense subgraph, then by applying Lemmas 2 and 3 we can embed a new subtree from the leftover of T.

2.2.6. The Second Mapping Procedure. Let Q denote the set of those vertices of $W-(M\cup D_1\cup D_2)$ which have not yet been covered by a vertex of T. In the Main Mapping Procedure we needed that every vertex in H has many neighbors in H. This time we allow Q to be very sparse, but assume that $|Q|>\gamma^{10}n$. The goal is to replace some of the covered vertices of G by those in Q in such a way that we gain a dense subgraph in the new Q.

Let $T'\subset T$ denote the portion of T that has already been embedded. Assume that T' has a subtree $\hat{T}$ with the following properties:

(i) $\hat{T}$ as a subtree of T' has exactly one neighbor in $T'-\hat{T}$;

(ii) $|\hat{T}|=t$, where $\gamma^{-20}\le t\le h_1=\gamma^{-40}\left(4K^3\gamma^{-10}\right)^{30(K+1)\log\gamma^{-1}}$;

(iii) The number of vertices of Q with degree at least $(1/2+\gamma^3)t$ into $\hat{T}$ is at least t^2.

If the above is satisfied, we call $\hat{T}$ a *remappable* subtree. Notice that if there are many edges from Q to a good subtree $\hat{T}$ such that $\hat{T}$ is connected to T' by one edge, then $\hat{T}$ is remappable.

Let $Z=\left\{z\in Q:\left|N(z)\cap I(\hat{T})\right|\ge(1/2+\gamma^3)t\right\}$, and construct the bipartite graph $F=F(A,B)$, where $A=I(\hat{T})$ and $B\subset Z$ such that $|B|=t^2$. We connect $a\in A$ and $b\in B$ by an edge if they are adjacent in G. We apply the first version of the Cleaning Lemma to find a subgraph $F'(A',B')\subset F(A,B)$ such that $A'\subset A$, $B'\subset B$, every $b\in B'$ has at least $(1+\gamma^3)t/2$ neighbors in A' and every $a\in A'$ has at least t neighbors in B'.

We apply Lemma 2 to find a mapping of $\hat{T}$ onto F' such that at most $t/2$ vertices of the new mapping cover vertices from $A=I(\hat{T})$ and the rest cover vertices from B. We re-embed $\hat{T}$ using vertices from M as in the Main Mapping Procedure, but we do not map a new subtree – seemingly it is a loss. On the other hand, there is a leftover bipartite graph $F''(A'',B'')\subset F(A,B)$, where $A''\subset A$ and $B''\subset B$, $|A''|\ge t/2$ and $|B''|\ge t^2-t$, and every vertex of B'' has at least γ^3t neighbors in A''. This dense bipartite subgraph can then be used to map a new unmapped subgraph of T by applying the Main Mapping Procedure. As is discussed in Remark 3, by Lemma 3 we are able to use F'' in order to map a subtree of size at least $\gamma^{-17}/2$. That is, we cover overall at most $2K^2$ vertices of M by the two applications of the Main Mapping Procedure, but map a new subtree which is of size at least $\gamma^{-17}/2$. The restrictions on the speed with which vertices of M are mapped are respected.

Good subtrees in the decomposition of $T - T_b$ may be connected to two other subtrees. Therefore we consider substituting (i) by

(i') $\hat{T} \subset T'$ has exactly two neighbors in $T' - \hat{T}$.

We show how to remap such a subtree. First, denote the two vertices of $\hat{T}$ which have neighbors in other subtrees by x and y, recalling that $x \neq y$. Denote the neighbors of x and y in the adjacent subtrees by z_1 and z_2, respectively. If the distance between x and y is at most three then we map the whole path connecting x and y onto vertices of M such that $I(x)$ is adjacent to $I(z_1)$ and $I(y)$ is adjacent to $I(z_2)$. This is easily done since M is a randomly chosen set. Then we map the subtrees of $\hat{T}$ below the vertices of the path by the Main Mapping procedure, using at most $6K^2$ vertices from M.

Otherwise denote the vertices of the path P connecting z_1 and z_2 by $z_1, x, \ldots, y'', y', y, z_2$. We map x and the part of $\hat{T}$ which is below P by the Main Mapping procedure, except the vertices y' and y and the subtrees which are below them. We map y'' to any vertex in Q. Since M is non-extremal, it is easy to find a path of length three connecting y'' and z_2 such the images of y' and y belong to M. Then we map the subtrees below y' and y by the Main Mapping procedure. Again, we use at most $6K^2$ vertices from M and as above we do not cover the vertices of M too quickly.

2.2.7. The Third Mapping Procedure. As in the second mapping method we assume that Q is very sparse and $|Q| > \gamma^{10}n$. Recall that the good subtrees of the decomposition are denoted by $\hat{T}_1, \hat{T}_2, \ldots, \hat{T}_\ell$. Let $t_i = v(\hat{T}_i)$. Assume that only a few vertices in Q have many neighbors in the good subtrees, and so we cannot apply the Second Mapping Method. More precisely, for every i there are less than t_i^2 vertices having at least $(1/2 + \gamma^3)t_i$ neighbors in $\hat{T}_i$.

We look for a *weakly remappable subtree*, that is, a good subtree $\hat{T}_j$ for which there are at least $\gamma^{20}n$ vertices in Q each having at least $(1/2 - \gamma^4)t_j$ neighbors in $\hat{T}_j$.

Assume that $\hat{T}_j$ is a weakly remappable subtree, and apply the second version of the Cleaning Lemma. We get a bipartite graph $F'(A', B')$ where $A' \subset V(\hat{T}_j)$, $B' \subset B \subset Q$, $|B| = t_j^3$, and every $a \in A'$ has at least t_j^2 neighbors in B' and every $b \in B'$ has at least $(1/2 - 2\gamma^4)t_j$ neighbors in A'.

If the sizes of the color classes of $\hat{T}_j$ differ by at least $\gamma^3 t_j$, then $\hat{T}_j \subset F'$. Moreover, after embedding $\hat{T}_j$ into F' we would get a leftover dense subgraph

of F' in which every b has at least $\gamma^3 t_j/2$ neighbors in the leftover of A'. This dense subgraph can then be used to build T further. Hence, if $\hat{T}_j$ is weakly remappable but we cannot proceed the way we described above, then the sizes of its color classes differ by at most $\gamma^3 t_j$.

Another important observation is that if $\hat{T}_j$ is weakly remappable, then we can embed most of it into F', a subtree $\widetilde{T}_j$ of size at most $2K\gamma^4 t_j$ will be left out. Assume that after embedding $\hat{T}_j - \widetilde{T}_j$ into F' the leftover of B' contains at least $t_j^2/2$ vertices, each having at least $\gamma^3 t_j$ neighbors in the leftover of A'. Then we are able to embed $\widetilde{T}_j$ into the leftover by the help of the Main Mapping Procedure. Moreover, since $v(\widetilde{T}_j) \leq 2K\gamma^4 t_j$, we can build T further by mapping a large subtree of it of size at least $\gamma^3 t_j/4$.

From now on we assume that none of the above conditions are satisfied, that is, the color classes of the weakly remappable subtree $\hat{T}_j$ are roughly of equal size, and apart from a few vertices, the neighbors of the vertices of B' are concentrated in a subset of A' of size $(1/2 - 2\gamma^4 - \gamma^3)t_j$.

Set $q = |Q|$. We show the following:

Lemma 12. *If Q does not have a subset Q' such that the induced subgraph on Q' has minimum degree at least $\gamma^{20}n$, then Q has at most $\gamma^{10}q$ vertices which have more than $\gamma^{10}n$ neighbors in Q.*

Proof. Applying Lemma 3, the number of edges inside Q cannot be more than $q\gamma^{20}n$. An easy calculation shows that the number of those vertices of Q which have at least $\gamma^{10}n$ neighbors in Q cannot be larger than $2\gamma^{10}q$. ■

We claim that most of the subtrees are weakly remappable if none of them are remappable. More precisely:

Lemma 13. *In the above setup if none of the good $\hat{T}_j$ subtrees are remappable then at least $(1 - \gamma/2)n$ vertices are in weakly remappable good subtrees.*

Proof. First observe that most vertices in Q have at most $\gamma^5 n/3$ neighbors which are not in the union of the $\hat{T}_i$. This follows from Lemma 12 and the fact that there are at most $\gamma^{10}n$ vertices which do not belong to good subtrees, $|D_1 \cup D_2 \cup M| < 3\gamma^8 n$ and T_b has size at most $\gamma^5 n/K^2$. Set $q'' = |Q''|$ where Q'' is the subset of Q containing those vertices having at most $\gamma^{10}n$ neighbors in Q.

Assume that $\hat{T}_j$ is not a weakly remappable subtree. Then there are at most $\gamma^{20}n$ vertices in Q'' which have more than $(1/2-\gamma^4)t_j$ neighbors in $\hat{T}_j$, with the remaining $q''-\gamma^{20}n$ vertices having fewer. Similarly, as $\hat{T}_i$ is not remappable, at most $\gamma^{20}n$ vertices have more than $(1/2+\gamma^3)t_i$ neighbors in $\hat{T}_i$. On the other hand, the number of edges going between Q'' and the union of the good subtrees is at least $q''n(1/2-\gamma^5/3)$.

Putting these together we get the following inequality:

$$\big(\gamma^{20}n+\big(q''-\gamma^{20}n\big)\big(1/2-\gamma^4\big)\big)c$$

$$+\big(\gamma^{20}n+\big(q''-\gamma^{20}n\big)\big(1/2+\gamma^3\big)\big)(n-c)\geq q''n\big(1/2-\gamma^5/3\big),$$

where c denotes the number of vertices in subtrees which are not weakly remappable. The dominating term on the left is $-q''\gamma^4 c$, while the dominating term on the right is $-q''n\gamma^5/3$. Therefore $c<\gamma n/2$, which implies the statement of the lemma. ■

Without loss of generality let $\{\hat{T}_1,\hat{T}_2,\dots,\hat{T}_k\}$ be the set of weakly remappable good subtrees. By the second version of the Cleaning Lemma there is a partial remapping of $\hat{T}_i$ which leaves a subtree of size at most $2K\gamma^4 t_i$ unmapped. Let $R_i\subset I(\hat{T}_i)$ denote the vertices in the image of $\hat{T}_i$ that are not used in the remapping. Let $r_i=|R_i|$. Observe that r_i is roughly half of t_i.

The plan again is to gain a large dense subgraph in Q, which we use to build T further. Let us assume that we cannot remap any of $\hat{T}_1,\hat{T}_2,\dots,\hat{T}_k$ such that a dense subgraph is left in Q. Then (R_i,Q) is sparse for every i, and $\sum r_i\geq n(1/2-2\gamma/3)$.

Consider the set $\cup R_i$. It has many edges, $e(\cup R_i)\geq\gamma n^2/3$, since G is non-extremal. We will show that there is a pair (R_i,R_j) which has many edges and with R_i and R_j roughly the same size.

Lemma 14. *There exists i, j such that $e(R_i,R_j)\geq\frac{\gamma}{2}r_ir_j$, moreover $r_i\leq r_j$ implies $r_j\gamma/4\leq r_i$.*

Proof. Let us assume that, on the contrary, every pair has density less than γ. Then

$$e(\cup R_i)\leq\gamma/4\sum_{i\neq j}r_ir_j+\sum_i e(R_i).$$

The sum $\sum_i e(R_i)=o(n^2)$, so we concentrate on the first sum, which is at most $\gamma/4\sum_i r_i(n-r_i)$. This last expression is at most $\gamma n^2/4-o(n^2)$. This contradiction implies that we can find a dense pair.

Assume now that all of the dense pairs are comprised of two sets with largely different sizes. Fix R_i and consider the pairs (R_i, R_j) for which $r_i < r_j$ – by assumption $r_j > 4r_i/\gamma$ if (R_i, R_j) is a dense pair. The number of edges in the dense pairs containing R_i is less than $\gamma r_i/4 \sum r_j \leq \gamma r_i n/8$. Summing over every i we find that strictly less than $\gamma n^2/3$ edges are in pairs comprised of sets of largely different sizes. This proves the lemma. ∎

We are now in a position to describe the re-mapping. If most of the subtrees are weakly remappable and none of them is remappable, then remap $\hat{T}_i, \hat{T}_j$ for which (R_i, R_j) is a dense pair as in Lemma 14. Assume that $t_i \leq t_j$. First remap most of $\hat{T}_i$, a subtree of size at most $2K\gamma^4 t_i$ will be left unmapped.

Since $\hat{T}_j$ is weakly remappable, we can find a bipartite graph $F(A', B')$ according to the second version of the Cleaning Lemma. When remapping most of $\hat{T}_i$ we may use vertices from B'. Notice, however, that apart from at most $t_j^2/2$ vertices, the vertices of B' have at most $\gamma^3 t_j$ neighbors in R_j. Moreover, every $a \in A'$ has at least t_j^2 neighbors in B'. Delete those vertices from B' which are used to remap $\hat{T}_j$ together with those which have more than $\gamma^3 t_j$ neighbors in R_j. We lose at most $t_j^2/2 + t_i$ vertices and so B' still has at least $2t_j^2/5$ vertices. Delete every edge going between R_j and B', so that every vertex of B' has at least $(1/2 - 2\gamma^3)t_j$ neighbors in A'. We can then remap most of $\hat{T}_j$ such that a subtree of size at most $2K\gamma^3 t_j$ is left out. In addition, we do not use any vertex of R_j during the remapping.

As (R_i, R_j) is a dense pair, there are at least $\gamma r_i r_j/2$ edges in between them. By Lemma 3 we can find a subgraph of this dense pair in which the minimum degree is at least $\gamma/2\frac{r_i r_j}{r_i + r_j}$. This is much larger than $2K\gamma^4 t_i$ or $2K\gamma^3 t_j$, since r_i, r_j and hence t_i and t_j does not differ by much. Therefore we can map the unmapped subtrees of $\hat{T}_i$ and $\hat{T}_j$, and still have most of the edges of the dense pair left to map a new, so far unmapped subtree of T. As usual, we ensure that the mapping is properly connected through the use of the random subset M by the help of the Main Mapping Procedure. Observe that since the sets of dense pairs have sizes at least $\gamma^{-20}/2$ we embed large new subtrees.

2.2.8. Description of the Embedding Algorithm. We are now ready to give the embedding method in the case where G is non-extremal and T has a broad subtree. After we are done with the mapping of the vertices of

$T_b - S$, we apply the decomposition method of Section 2.2.2. Then we map the good subtrees of the decomposition in a top-down fashion.

Starting from the root r, we map large subtrees of T one-by-one. During the embedding the already mapped part of T will be a tree $T' \subset T$, that is, it will be connected. We proceed in the following way. We pick a leaf x of T' such that there is a subtree $T_x \subset T(x)$ of size $\gamma^{-12} < v(T_x) < \gamma^3 h_1$. We want to extend I for $T' \cup T_x$.

As before, let Q denote the set of uncovered vertices of $W-(M\cup D_1\cup D_2)$. If Q has a large dense subgraph, then T_x can be mapped by the Main Mapping Procedure, recall Remark 3. If not, then we will rethink the way we mapped certain subtrees, and look for remappable subtrees. If Q is not too small there will always be a remappable subtree of size at least γ^{-20}. Remapping this subtree returns a large dense bipartite subgraph to Q which can be used to map T_x.

If there is no remappable subtree then most of the already embedded subtrees are weakly remappable by Lemma 13. There exist two good subtrees which, upon remapping, return a large dense pair to Q (Lemma 14). This dense pair can be used to map T_x.

We proceed in this way until the size of Q drops below $\gamma^{10}n$. Then we apply the matching method of Section 2.2.4, and find an embedding of T into G. This finishes the proof in the first non-extremal case. ■

2.3. The second non-extremal case: T has a long subtree

In this section we discuss the case of embedding T when there is a subtree $T_l \subset T$ with at least $\gamma^5 n/K^3$ vertices and at most $\gamma^7 n$ leaves. We first give a sketch of the procedure.

A *line* in a tree T is a path P such that every vertex on P apart from possibly the endpoints has degree 2 in T. In this case, T_l must contain several long lines. We find a set of $2\gamma^{10}n$ lines, each of length $1/(5\gamma^2)$. We will cut out the midpoint from each of these lines, set aside the cut-out vertex, and glue the two new endpoints together. The result is a tree $\widehat{T}$ on $(1-2\gamma^{10})n$ vertices.

We prepare G for the embedding of $\widehat{T}$, beginning with finding the vertices to be covered by the long lines by the help of a randomized procedure. Then we apply the embedding method of the previous section for embedding $\widehat{T}$ into G. Since $\widehat{T}$ is smaller than G, the embedding goes easily, even with

the restriction that the long lines of $\widehat{T}$ are mapped onto the pre-determined long paths of G.

Finally we address the cut-out vertices. We will extend the embedding of the long lines by finding a perfect matching in an appropriately defined bipartite graph in G. With high probability, this procedure finds an embedding of T into G.

2.3.1. Preparations for the embedding

Finding random paths in G. In the first step we pick two random subsets $M_1, M_2 \subset W(G)$. M_1 will have size $\gamma^{10}n$, while M_2 will contain $\gamma^8 n$ vertices. Then we find the paths U_i by the following randomized procedure:

Set $k = 2\gamma^9 n$. Pick randomly and independently $\frac{k}{20\gamma^2}$ edges from $G - M_1 - M_2$ with replacement. From these, form k sets of size $1/(20\gamma^2)$. We can connect the edges of each subset into a path of length $\frac{1}{5\gamma^2}$ by way of the paths of length three guaranteed by Lemma 8. Denote the paths by $U_1, U_2, \ldots, U_k$. We discard those paths containing repeated vertices, where a vertex is repeated if either it appears twice in some U_i or it is contained by U_i and U_j for $i \neq j$.

Let v be any vertex in G. Let X_v denote the number of occurrences of v in the randomly chosen edges, then $\mathrm{E}X_v < \gamma^7$. This number follows a Poisson distribution,

$$\Pr(X_v > 2) \leq \sum_{i \geq 2} \frac{\gamma^{7i} e^{-\gamma^7}}{i!} \ll \gamma^{13}.$$

That is, on average we expect to have less than $\gamma^{13}n$ repeated vertices. The probability that there are more than $10\gamma^{13}n$ repeated vertices is by Markov's inequality at most $1/10$. We call a path *bad* if either it contains a vertex at least twice, or it contains a vertex which appears in another path. In the latter case we call only one of those a bad path. After discarding the bad paths, the vast majority of the paths remain. With probability at least 90% we have at most $\frac{10\gamma^{13}n}{8\gamma^2} < 2\gamma^{11}n$ bad paths.

A vertex v can be inserted into a path U_i if v has two consecutive neighbors on the path. We also say that v is adjacent to U_i. Let us estimate the probability that v is not adjacent to a given path U_i. There are at least γn^2 edges in the neighborhood of v, since G is non-extremal. The probability

that none of the $\frac{1}{20\gamma^2}$ randomly chosen edges is in the neighborhood of v is at most

$$(1-\gamma)^{\frac{1}{20\gamma^2}} \leq e^{-\frac{1}{20\gamma}} < \gamma^{20}.$$

It is easy to upper bound the number of those paths for which there are less than $(1-\gamma^{11})n$ adjacent vertices. By Markov's inequality, the probability that one cannot insert more than $(1-\gamma^{11})n$ vertices into a path is γ^9. Using Markov's inequality again the number of such paths is at most $\gamma^{10}n$ with probability $\geq 1-\gamma/3$.

Hence, one can find $k' = \gamma^9 n$ paths such that these do not contain repeated vertices, and one can insert at least $(1-\gamma^{11})n$ vertices into any of them.

Finding long lines in T. First we show that if F is a tree having only a few leaves then it contains many long lines.

Lemma 15. *Let F be a tree on t vertices with ct leaves for some $0 < c \ll 1$. Then it is possible to find $s = ct$ vertex disjoint lines $p_1, \ldots, p_s$ in F such that $|p_i| = 1/(4c)$.*

Proof. Choose a root ρ. Substitute every maximal line in F by one edge (but we keep ρ as is). Then every vertex in the new tree F' will have degree 1 or at least 3 except possibly the root. Since in a tree the number of leaves is bounded by the number of vertices with degree at least 3 plus 2, we have at most $2ct+1$ vertices and $2ct$ edges in F'. There were $t-1$ edges in F, hence an average edge of F' corresponds to a line of length a least $1/(2c)$. Cut out a part of this line of length $1/(4c)$ and glue together the resulting two endpoints. Repeat this procedure: construct a new F' as before, and then find a long path again by an averaging argument. We can continue this way, and find many paths of length $1/(4c)$, as long as the leftover number of edges in F is larger than $t/2$. ■

Apply the decomposition of Section 2.2.2 for T_l and for the rest of T separately. Call a good subtree of T_l *long* if it has size t and has at most $2\gamma^2 t/K^3$ leaves. It is easy to see that almost half of the vertices of T_l are in long good subtrees. Apply Lemma 15 to find k' lines of length $1/(4\gamma^2)$. We return edges of the lines to T as necessary in order to ensure that any two are at a distance of at least three from each other and that they are exactly $1/(5\gamma^2)$ long. We denote the resulting lines by $P_1, P_2, \ldots, P_{k'}$. Notice, that we may use up more than 40% of a long good subtree when determining the long lines. Therefore the long subtrees we use for the long lines will not contain more than $3\gamma^7 n$ vertices overall.

Let P_i be a long line. Let y_i be the midpoint of P_i and x_i and z_i its neighbors. Remove y_i from the path, and include the x_iz_i edge. This yields a new line $\widehat{P}_i$ and the cut-out vertex y_i. We repeat this for every $1 \leq i \leq k'$, after which we arrive at a tree $\widehat{T}$ which has $(1-k')n$ vertices.

2.3.2. Embedding $\widehat{T}$ into G. Our next goal is to embed $\widehat{T}$ into G. The decomposition of T carries over to one for $\widehat{T}$, comprised of good subtrees plus a small leftover. We use the embedding method of the previous case with minor modifications. In this case there is no use of D_1 and D_2 as their role is replaced by the random embedding of the long lines and the removal of one vertex from each line, which we embed at the end. The lines P_i are embedded first into the randomly constructed paths in G arbitrarily. The rest of the embedding of the decomposed $\widehat{T}$ proceeds very similarly as before, with the exception that we begin with embedding T_l, the long subtree, and when mapping a good subtree of T_l containing the line $\widehat{P}_i$ we apply the Main Mapping Procedure with the random set M_2, i.e. connecting through M_2, in order to connect the already embedded line to the rest of the subtree. Since overall we have $\gamma^9 n$ such lines, and we use at most $6K^2$ vertices to connect the embedded line, we do not use the vertices of M_2 too quickly. After we are done with T_l, we put back the uncovered vertices of M_2 into the set of vacant vertices of $W - M_1$. Note that from this point the embedding procedure of the previous section is able to proceed as long as there are enough vertices in G that are uncovered. As we have retained $k' \gg |M_1|$ vertices to embed at the very end, this condition is always maintained.

2.3.3. Finishing the embedding. Recall that the probability that a vertex v cannot be inserted into any of the $\frac{1}{20\gamma^2}$ randomly chosen edges is at most

$$(1-\gamma)^{\frac{1}{20\gamma^2}} \approx e^{-\frac{1}{20\gamma}}.$$

Since we choose the edges and then form the edge sets randomly, by Chernoff's inequality every vertex v can be inserted to at least $(1-e^{-\frac{1}{20\gamma}})\gamma^9 n - o(n)$ paths with probability $\geq 1 - 1/n^2$.

After embedding $\widehat{T}$, we have the uncovered $A \subset W(G)$ such that $|A| = k'$. If we can insert exactly one vertex of A to each of the U_i paths, then we are done with the embedding of T into G. For showing that these insertions are possible we will apply the König–Hall marriage theorem.

Construct a bipartite graph $F(A, B)$ where A is as above and the vertices of B correspond to the randomly chosen U_i paths. We connect $a \in A$ and

$b \in B$ if the corresponding vertex can be inserted to the the path of b. We have $\gamma^9 n$ vertices in A and B. Recall that we discarded those paths into which less than $(1-\gamma^{11})n$ vertices could be inserted. Furthermore, with high probability every vertex in A is adjacent to most of the paths, thus, the minimum degree in F is large with high probability. Hence, the König–Hall conditions are satisfied with high probability. Therefore, we can find a perfect matching in F. This finishes the proof of the second non-extremal case. ■

3. The Extremal Case

We demonstrate how to embed a tree T into an extremal graph G. G is extremal if it is very close to being either a complete bipartite graph on color classes having size $n/2$, or it is the union of two complete graphs on $n/2$ vertices each. The two cases, while very similar, are handled separately.

Note that that we did not use the full strength of the minimum degree of G in the non-extremal case. Even a minimum degree of $(1/2-\varepsilon)n$ for some small ε is sufficient for the arguments in the non-extremal case. The degree condition is critical, however, in the extremal case. If T has unequal color classes and $G = K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, then $T \not\subset G$. If $G = K_{n/2} \cup K_{n/2}$, then it is not even connected. The First Extremal Case is when G is close to $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$, the Second Extremal Case is when G is close to $K_{n/2} \cup K_{n/2}$.

We assume that $\delta(G) \geq n/2 + c_K \log_2 n$, where c_K is a constant depending only on K (for more we refer to Lemma 17 and Lemma 18 below). In both extremal cases, the proof relies on this minimum degree. In the last section we will show that $\delta(G) - n/2 = \Omega(\log n)$ in order to contain every bounded degree spanning tree.

We will make use of a folklore result concerning trees in both extremal cases.

Proposition 16. *Let J be any tree on m vertices. Then J has a vertex $x \in V(J)$ such that it is possible to group the vertices of $J - x$ into two forests, J_1 and J_2 such that $m/3 \leq v(J_1), v(J_2) (\leq 2m/3)$ and there is no edge connecting J_1 and J_2 in $J - x$.*

The vertex x will be called a *split vertex*. Observe, that Proposition 16 applies for forests as well. If J is a forest, then turn it into a tree by adding

edges to it. One can apply the proposition, find x, J_1 and J_2, and then delete the recently added edges.

In one of the extreme cases we have to cut a tree into two sets of equal size such that there is only a few edges going in between the two sets.

Lemma 17. *Let J be a tree on n vertices for some even integer n. J can be divided into two sets of vertex disjoint subtrees $J_1, J_2, \ldots, J_t$ and $J_{t+1}, \ldots, J_s$ such that the number of vertices in each sets is $n/2$, and $s \leq 2\log_2 n$. If the maximum degree in J is D, then the number of edges between the two set of subtrees is at most $2D\log_2 n$.*

Proof. Let $\mathcal{A}$ and $\mathcal{B}$ be two empty sets. Each time we split a subforest of J, we get a smaller forest and a larger forest. We continue to split on the larger forest until it can fit into $\mathcal{A}$, then repeat the process on the smaller forest from the last splitting. The process is iterated until we fill $\mathcal{A}$ perfectly. After $O(\log n)$ splitting steps we get that $|\mathcal{A}| = |\mathcal{B}| = n/2$.

We define the a_i numbers for $1 \leq i \leq O(\log n)$, a_i denotes the empty space left in $\mathcal{A}$ at the ith step. In particular, $a_0 = n/2$. In the first step we apply Proposition 16 and remove the vertex x, splitting J into J_{in} and J_{out} such that $n/3 \leq v(J_{\text{in}}) \leq v(J_{\text{out}})$. We put J_{in} into $\mathcal{A}$, and set $a_1 = a_0 - v(J_{\text{in}})$ from which it follows that $a_1 \leq a_0/2$. Define i_1 so that

$$\frac{n}{2^{i_1}} \leq a_1 < \frac{n}{2^{i_1-1}}.$$

Define $J^0 = J_{\text{out}}$, and form a sequence of forests $J^0, J^1, \ldots J^{s_1}$ by recursively splitting the larger forest from the previous iteration, i.e. J^l is the larger forest arising from splitting J^{l-1}. Define s_1 such that $v(J^{s_1}) < a_1 < v(J^{s_1-1})$ and let $J^{s_1-1} = J^{s_1}_{\text{in}} + J^{s_1}_{\text{out}} + x$, where x is the splitting vertex chosen from J^{s_1-1} and where $v(J^{s_1}_{\text{out}}) < v(J^{s_1}_{\text{in}})$. Since after two applications of splitting the size of the larger forest is less than half of that of the original, we have that $s_1 \leq 2i_1$. We put the vertices of $J^{s_1}_{\text{in}}$ into $\mathcal{A}$, let $a_2 = a_1 - v(J^{s_1}_{\text{in}})$. By the above conditions $a_2 \leq a_1/2$ and $a_2 \leq v(J^{s_1}_{\text{out}})$. If equality holds in the latter inequality, put the vertices of $J^{s_1}_{\text{out}}$ into $\mathcal{A}$ and stop. Let $J^{s_1} = J^{s_1}_{\text{out}}$. This is the first step of the process.

In general, assume that we have performed $k-1$ steps and that there are a_k vertices left to be placed in $\mathcal{A}$, where

$$\frac{n}{2^{i_k}} \leq a_k < \frac{n}{2^{i_k-1}}$$

for some positive integer i_k. Let $S_{k-1} = s_1 + s_2 + \dots s_{k-1}$, and assume that $s_t \leq 2(i_t - i_{t-1})$ for $2 \leq t < k$. The forest $J^{S_{k-1}}$ is to be split further, where $a_k < v\big(J^{S_{k-1}}\big) < n/2^{i_{k-1}-1}$.

Splitting on the larger subforest as above $s_k \leq 2(i_k - i_{k-1})$ times we have J^{S_k}, where $S_k = S_{k-1} + s_k$ and

$$v\big(J^{S_k}\big) \leq a_1 < v\big(J^{S_k-1}\big).$$

Define

$$J^{S_k-1} = J_{\text{in}}^{S_k} + J_{\text{out}}^{S_k} + x$$

where x is the splitting vertex chosen from J^{S_k-1} and

$$v\big(J_{\text{out}}^{S_k}\big) \leq v\big(J_{\text{in}}^{S_k}\big).$$

We put the vertices of $J_{\text{in}}^{S_k}$ into $\mathcal{A}$, let $a_{k+1} = a_k - v\big(J_{\text{in}}^{S_k}\big)$ and let $J^{S_k} = J_{\text{out}}^{S_k}$. It is easy to see that $a_{k+1} \leq a_k/2$ and $a_{k+1} \leq v\big(J^{S_k}\big)$. If equality holds in the latter inequality, put the vertices of J^{S_k} into $\mathcal{A}$ and stop.

Clearly, since at every step we halve the space left to fill in $\mathcal{A}$, we need at most $\log_2 n$ steps in order to get that $|\mathcal{A}| = n/2$. The vertices of those forests which were not put into $\mathcal{A}$ are put into $\mathcal{B}$. Since

$$S_k \leq 2i_1 + \sum_{1<t<k} 2(i_t - i_{t-1}) = 2i_{k-1}$$

holds for every k we get that we have to apply Proposition 16 at most $2\log_2 n$ times. From this the lemma follows easily. ■

Remark 4. One can think of the above splitting procedure as giving labels to the vertices of T. If $x \in V(T)$, then its label $\lambda(x)$ is either A or B. The number of those xy edges for which $\lambda(x) \neq \lambda(y)$ is $O(\log n)$, since in such a case either x or y has to be a split vertex.

In the other extremal case we have to find an almost proper 2-coloring of the vertices of a tree on n vertices, such that there are at most $O(\log n)$ edges which connect two vertices with the same color.

Lemma 18. *Let J be a tree on n vertices, n being an even integer. J can be divided into vertex disjoint subtrees $J_1, J_2, \dots, J_s$ plus at most $2\log_2 n$ split vertices, such that there is a proper 2-coloring of the vertices of the subtrees with equal sized color classes.*

Proof. As in the proof of Lemma 17 our main tool will be Proposition 16. The goal is to give an "almost proper 2-coloration" of the tree J. The color classes will be of the same size, and apart from $O(\log n)$ edges, every edge will connect vertices with different colors.

For a tree F let us define the *coloring discrepancy*, $d(F)$ to be the difference of the sizes of the color classes. We denote by χ_i, $i = 1, 2$ the size of color class i, so that $d(F) = \big|\chi_1(F) - \chi_2(F)\big|$. If F has m vertices then $d(F) < m$. If we have l trees, $F_1, F_2, \ldots, F_l$, each having size at most m, then the forest $\cup F_i$ has a proper 2-coloration such that $d\big(\bigcup_i F_i\big) < m$.

We exhibit a coloring algorithm which colors every vertex of J in $O(\log n)$ steps. We assume that there is a proper 2-coloring of J in the beginning, and the algorithm re-colors the vertices such that the color classes are of equal size and at most $O(\log n)$ edges connect vertices of the same color.

Let $\widetilde{\alpha}_i$ be the number of vertices whose color has been reassigned to the 1st color class in the ith step, and let $\widetilde{\beta}_i$ be the number that have been reassigned to the 2nd color class. Let $\alpha_i = n/2 - \widetilde{\alpha}_i$ and $\beta_i = n/2 - \widetilde{\beta}_i$. Before any recoloring, $\widetilde{\alpha}_0 = \widetilde{\beta}_0 = 0$ and $\alpha_0 = \beta_0 = n/2$. Throughout we will assume that $\alpha_i \leq \beta_i$, swapping the two colors at the end of the $i-1$st step if necessary. We let $d^i_{\text{in}} = \beta_i - \alpha_i \geq 0$, representing the coloring discrepancy of the re-colored vertices of J. At the ith step there is a forest J^i_{out} containing those vertices which have not been re-colored, the "outside forest". We denote its coloring discrepancy by d^i_{out}, determined by the original 2-coloring of J restricted to J^i_{out}. Note that $\alpha_i + \beta_i = v(J^i_{\text{out}})$. If J^i_{out} is not a tree, then one can color it in many different ways.

We will maintain the following two conditions throughout the algorithm:

C1: in the ith step $v(J^i_{\text{out}}) \leq n/2^i$ and

C2: $d^i_{\text{in}} \leq d^i_{\text{out}}$.

These conditions ensure that d^i_{in} is decreasing fast, since $d^i_{\text{in}} \leq d^i_{\text{out}} \leq v(J^i_{\text{out}}) \leq n/2^i$. Therefore, after $O(\log n)$ steps we can get the desired coloring of J. We demonstrate how to achieve $C1$ and $C2$ in the first step and how to maintain these conditions throughout the procedure, implying the lemma.

To begin, we apply Proposition 16 twice. After the first application we get two forests, and we apply splitting once more on the larger forest. One of the split vertices are put into the 1st color class, the other to the 2nd color

class. We will have the three sub-forests J_1, J_2 and J_3 such that $v(J_i) \leq n/2$ for $i = 1, 2, 3$. Assume that $d(J_1) \leq d(J_2) \leq d(J_3)$, where the discrepancy is taken on the original coloring of J restricted to these sub-forests. Then we recolor J_1 and J_2 such that $d(J_1 \cup J_2) \leq d(J_3)$. This is easily accomplished by swapping the colors in J_1, say, so that the discrepancy of the recolored union is $d(J_2) - d(J_1) < d(J_3)$. We let $d^1_{\text{in}} = d(J_2) - d(J_1)$ and $J^1_{\text{out}} = J_3$. Conditions C1 and C2 are clearly satisfied.

Now assume that we have completed the ith step, $v(J^i_{\text{out}}) \leq n/2^i$ and $d^i_{\text{in}} \leq d^i_{\text{out}}$. As before we apply Proposition 16 twice. First we split J^i_{out} into two forests, then split the larger one, yielding the forests J_1, J_2 and J_3. One of the split vertices are put into the 1st color class, the other to the 2nd color class. We have that $v(J_t) \leq n/2^{i+1}$ for $t = 1, 2, 3$. The sizes of the color classes of J_t will be denoted by a_t and b_t, where $a_t \leq b_t$ for $t = 1, 2, 3$. Observe that $\sum(a_t + b_t) = \alpha_i + \beta_i$.

Claim 19. $\sum_{1 \leq t \leq 3} a_t \leq \alpha_i$.

Proof (of the claim). Suppose that $\sum a_t > \alpha_i$, then $\sum b_t < \beta_i$. Therefore,

$$d^i_{\text{out}} = \sum b_t - \sum a_t < \sum b_t - \alpha_i < \beta_i - \alpha_i = d^i_{\text{in}},$$

contradicting condition C2. ∎

A forest J_t does not fit into the leftover, i.e., it cannot be re-colored without further splitting, if either $a_t > \alpha_i$ or $b_t > \beta_i$. We have just eliminated the first option. We say that J_t cannot be re-colored if $b_t > \beta_i$.

Claim 20. *There is at most one forest among J_1, J_2 and J_3 which cannot be re-colored.*

Proof (of the claim). Assume that there are two forests which cannot be re-colored, say, J_1 and J_2. Then $b_1, b_2 > \beta_i$, hence,

$$\sum(a_t + b_t) \geq b_1 + b_2 > 2\beta_i \geq \beta_i + \alpha_i = \sum(a_t + b_t),$$

a contradiction. ∎

Assume that $b_1, b_2, b_3 \leq \beta_i$, that is, each of the three forests can be re-colored. Further assume that $d(J_1) \leq d(J_2) \leq d(J_3)$. Then we first exchange the colors on the vertices of J_1 so that the new coloring discrepancy is $\left|d(J_1) - d^i_{\text{in}}\right|$, easily seen to be at most $d(J_2) + d(J_3)$. Next we exchange

the colors of the vertices of J_2, which changes the coloring discrepancy to $d_{\text{in}}^{i+1} = \left|d(J_2) - \left|d(J_1) - d_{\text{in}}^{i}\right|\right| \leq d(J_3)$. We let $J_{\text{out}}^{i+1} = J_3$ so that $v(J_{\text{out}}^{i+1}) \leq n/2^{i+1}$. Hence, conditions C1 and C2 are satisfied.

Now assume that $b_3 > \beta_i$ and $b_1, b_2 \leq \beta_i$. We claim that in this case $d(J_3) > d(J_1), d(J_2)$. If this were not true, then say $d(J_1) > d(J_3)$. Clearly

$$2\beta_i \geq \alpha_i + \beta_i \geq b_1 + a_1 + b_3 + a_3,$$

and so

$$2\beta_i - b_3 \geq b_1 + a_1 + a_3,$$

implying that

$$2\beta_i - b_3 - 2a_1 \geq a_3 + b_1 - a_1.$$

From the assumption that $d(J_1) > d(J_3)$ we get that $b_3 < a_3 + b_1 - a_1$, implying that

$$b_3 < 2\beta_i - b_3 - 2a_1.$$

This implies that $b_3 < \beta_i - a_1$, which is a contradiction.

In short, either all the three forests can be re-colored, or the one which cannot has the largest coloring discrepancy. In the latter case we can repeat what we did in the first step by first re-coloring J_1, then J_2. We let $J_{\text{out}}^{i+1} = J_3$. As above, we have $d_{\text{out}}^{i+1} \geq d_{\text{in}}^{i+1}$ and $v(J_{\text{out}}^{i+1}) \leq n/2^{i+1}$. Again, conditions C1 and C2 are satisfied. In at most $\log_2 n$ steps we can find the desired 2-coloring. Since at every step we apply Proposition 16 twice, the total number of split vertices is at most $2\log_2 n$, and this finishes the proof of the lemma. ■

Remark 5. As in the previous case, one can think of the above splitting procedure as giving labels to vertices of T. If $x \in V(T)$ then $\lambda(x)$ is either A or B, that is, the label of a vertex is a set. The number of those xy edges for which $\lambda(x) = \lambda(y)$ is $O(\log n)$, since in such a case either x or y has to be a split vertex.

The following embedding lemma will be applied in both type of extremal cases:

Lemma 21. *Let F be a tree with constant maximum degree K and color classes Q_F and R_F. Let $H = (Q_H, R_H)$ be a bipartite graph such that*

1. $q = |Q_F| = |Q_H|$ *and* $r = |R_F| = |R_H|$ *and* $m = q + r$;

2. *if* $u \in Q_H$ *then* $\deg(u) \geq \left(1 - \frac{1}{3K}\right)r$;

3. *if* $v \in R_H$ *then* $\deg(v) \geq \left(1 - \frac{1}{3K}\right)q$;

4. $Q \subset Q_F$ *and* $R \subset R_F$ *are such that* $|Q| + |R| = O(\log m)$,

where Q *and* R *denote the set of those vertices of* F *which are a priori restricted to be mapped onto given vertices of* Q_H, *respectively* R_H. *Assume further that* $e_F(Q, R) = 0$. *Then* F *is a spanning tree of* H *if* m *is sufficiently large.*

Proof. We will find a bijective, adjacency preserving mapping ϕ from $V(F)$ to $V(H)$. Note that $\phi(Q)$ and $\phi(R)$ are already given. First we map the vertices of $N_F(Q)$. If $x \in N_F(Q)$ then we pick a vacant vertex $v \in R_H - \phi(R)$ such that if $y \in N_F(x) \cap Q$ then v is adjacent to $\phi(y)$. We have at least $2r/3 - O(\log m)$ vertices to choose from since the minimum degree is large in H and we have to map $O(\log m)$ vertices. Then we let $\phi(x) = v$. Let $R = R \cup N_F(Q)$. Clearly, $|R| = O(\log m)$ since $\Delta(F)$ is a constant.

We extend the definition of ϕ, beginning with a random bijective mapping $\phi' : Q_F - Q \to Q_H - \phi(Q)$, mapping every $x \in Q_F$ onto $\phi'(x) \in Q_H$. Let $\phi(x) = \phi'(x)$ for every $x \in Q_F - Q$. We estimate the probability that the vertices of $N_F(R)$ are mapped onto neighbors of $\phi(R)$. We call ϕ *good* for R if $x \in N_F(R)$ and $y \in N_F(x) \cap R$ implies $\phi(x)\phi(y) \in E(H)$.

Let $x \in N_F(R)$ and assume that $y_1, y_2, \ldots, y_t \in R$ are adjacent to x, where $1 \leq t \leq K$. By the minimum degree condition of H we get that $\left|\cap_i N(y_i)\right| \geq 2q/3 - O(\log m)$, and so the probability of choosing a good $\phi'(x)$ from the uncovered vertices in Q_H is at least $1/2$ for every $x \in N_F(R)$. This implies that the probability that ϕ is good for R is at least $2^{-K|R|} = 1/\operatorname{poly}(m)$, where $\operatorname{poly}(m)$ is some polynomial of m.

We extend ϕ to an embedding of F by a matching argument. Let $R'_F = R_F - R$ and $R'_H = R_H - \phi(R)$. Construct the auxiliary bipartite graph $L(R'_F, R'_H)$, connecting $u \in R'_H$ to $x \in R'_F$ iff u is adjacent to $\phi(y_1), \phi(y_2), \ldots, \phi(y_K)$ in H, where $y_1, y_2, \ldots, y_K \in Q_F$ are the neighbors of x. If we can find a perfect matching in L then we have extended ϕ to an embedding of F into H.

For proving the existence of the perfect matching we check the König–Hall conditions. In particular, we show that every $u \in R'_H$ is adjacent to at least $r/2$ vertices in R'_F and that every $x \in R'_F$ is adjacent to at least $r/2$ vertices in R'_H with high probability.

Let $u \in R'_H$. The probability that $ux \in E(L)$ for some $x \in R'_F$ is at least

$$\frac{\binom{(1-\frac{1}{2K})q}{K} K!(q-K)!}{q!} \approx \left(1 - \frac{1}{2K}\right)^K > 0.56,$$

as this is the probability that the neighbors of x are all mapped to $N_H(u)$. By Azuma's inequality (see e.g., in [1]) we get that every $u \in R'_H$ is adjacent to more than $r/2$ vertices of R'_F with very high probability. In fact the probability that the degree of u in L is less than $r/2$ is at most $2e^{-cm}$ for some $c > 0$. Hence, the probability that ϕ is good for R and every vertex of R'_H has at least $r/2$ neighbors in L is $\geq (1 - 2e^{-cm}) - \big(1 - 1/\operatorname{poly}(m)\big) > 1/\big(2\operatorname{poly}(m)\big)$ if m is sufficiently large.

Next we show that every $x \in R'_F$ has at least $r/2$ neighbors in R'_H. Let the neighbors of $x \in R_F$ be $y_1, y_2, \ldots, y_K \in Q_F$. By the minimum degree condition in H we have that $\phi(y_1), \phi(y_2), \ldots, \phi(y_K)$ has at least $2r/3 - O(\log m) > r/2$ common neighbors in R'_H. Indeed, this holds for every ϕ. This proves the lemma. ■

Sketch of the embedding algorithm. Denote by SP the induced subforest of T which contains the split vertices and those vertices of T which are at a distance at most four from the split vertices. Clearly, $v(SP) \leq 2K^4 \log_2 n$. SP may contain many connected components.

In both extremal cases our first goal is to embed the forest SP in such a way that those vertices of SP which are adjacent to some vertex in $V(T) - V(SP)$ are mapped onto "large degree vertices" of G. In the first extreme case the large degree vertices are adjacent to at least $n\big(1/2 - 1/(4K)\big)$ vertices inside the set to which they belong (either in A or in B). In the second extreme case large degree vertices are those with at least $n\big(1/2 - 1/(4K)\big)$ neighbors in the other set. Since $v(SP) = O(\log n)$, after some preparations we can apply Lemma 21 in order to map the vertices of $T - SP$.

3.1. The First Extremal Case

In this case $W(G)$ is partitioned into A and B such that $|A| = |B| = n/2$, and $e\big(G|_A\big), e\big(G|_B\big) \geq \binom{n/2}{2} - \gamma n^2$.

First we state a simple lemma on the degrees of the vertices of A and B, we leave the proof to the reader.

Lemma 22. *Let $0 < \eta < 1/2$. Then the number of vertices of A having less than $n/2 - \eta n$ neighbors in A is at most $\frac{2\gamma n}{\eta}$. The analogous for B also holds.* ■

Preparations for the embedding – Step 1: Decomposition of T. By Lemma 17 we find a bipartition of the vertex set of T into two sets, X and Y such that $|X| = |Y| = n/2$ and $e(X,Y) = O(\log n)$. Recall that SP is the sub-forest of T which contains the split vertices and those vertices of T that are at a distance of at most four from the split vertices. Let $SP_1, SP_2, \ldots, SP_t$ denote the components of SP. Since SP is an induced subgraph, there are no edges connecting SP_i and SP_j for $i \neq j$. For every $1 \leq i \leq t$ we pick an arbitrary split vertex belonging to SP_i to be the root of SP_i.

Preparations for the embedding – Step 2: Switching and finding stars. We define the *inner degree* of $v \in A$ as $\left|N_G(v) \cap A\right|$ and the *outer degree* of $v \in A$ as $\left|N_G(v) \cap B\right|$. The inner and outer degree of $u \in B$ is defined similarly.

Assume that $u \in A$ and $v \in B$ are such that

$$e(A - u + v) + e(B - v + u) > e(A) + e(B).$$

Switch u to B and v to A, and look for another pair of vertices which can be switched. At every switching step the total number of edges inside A and B is increased, therefore this process will stop in a finite number of steps. Observe that if there exists $u \in A$ and $v \in B$ such that both have inner degrees of at most $n/4$, then it is still possible to perform a switching step. Hence, at the end of the switching procedure we have that every vertex in one of the parts, say A, has an inner degree of at least $n/4$. Observe, that after the switching even the sparser set will contain at least $n^2/4 - 2\gamma n^2$ edges, so both sets will be 2γ-extremal. There may be up to $16\gamma n$ vertices in B which have at most $n/4$ neighbors in B (see Lemma 22). Denote the set of these vertices with low degree in B by B'.

We determine $|B'|$ vertex disjoint stars in B, each star centered on a vertex of B' connected to K vertices from $B \setminus B'$.

We determine the stars greedily. For each $u \in B'$, assign K vertices from $N(u) \cap (B \setminus B')$ as long they are not yet assigned to another vertex of B'. Call a vertex of B' *marked* if it is designated as the center of a K-star, and call a vertex in $B \setminus B'$ *vacant* if it is not assigned to a marked vertex of B'.

The greedy procedure stops when every unmarked vertex of B' has less than K vacant neighbors in $B \setminus B'$.

If no unmarked vertex in B' has K vacant neighbors, then apart from at most $16K\gamma n$ vacant vertices the rest of $B - B'$ has no neighbor among the unmarked vertices of B'. Denote the set of unmarked vertices of B' by B'_u. B has $n/2$ vertices, of which at most $(K+1)16\gamma n$ belong to the stars and at most $(K+1)16\gamma n$ vertices have a neighbor in B'_u. Each of the remaining at least $n/2 - 32(K+1)\gamma n$ vertices has at least $c_K \log n$ neighbors in A. It follows that an average vertex of A has at least

$$c_K \log n \frac{n/2 - 32(K+1)\gamma n}{n/2} > K$$

neighbors among these vacant vertices. Hence, there is at least one $w \in A$ which has at least K vacant neighbors. We pick an arbitrary vertex $v \in B'_u$, switch it with w, mark w, and assign the K vacant neighbors to w. Note that since the greedy algorithm could not assign K vacant neighbors to v, the outer degree of v before switching was large, at least $n/2 - 16(K+1)\gamma n$.

The simple averaging argument ensures that there always will be vertex in A having K vacant neighbors, and so we can continue the process to guarantee that (1) every vertex of A has inner degree at least $n/4$, and either (2) the inner degree of a vertex $u \in B$ is at least $n/4$, or (3) u sits in the center of a K-star such that the leaves of this star have inner degree at least $n/4$.

Preparations for the embedding – Step 3: Embedding SP. It is easy to see that one can embed the induced sub-forest SP greedily if $c_K \geq 2K^4$. Still, we have to be careful to ensure that certain vertices of SP are mapped onto vertices with large inner degree.

Let $x \in V(SP)$, and denote its neighbors by $y_1, y_2, \ldots, y_K$. If x is the vertex to be mapped and some of the y_is have already been mapped, then we must map x onto an uncovered vertex $u \in \lambda(x)$ such that $uI(y_i) \in E(G)$ for every mapped y_i. We begin mapping the components of SP with their roots. We obey the following rules when mapping a neighbor of an already mapped vertex:

R1. If x is mapped to a vertex with inner degree at least $n/4$, y_i is unmapped and $\lambda(x) = \lambda(y_i)$, then we map y_i onto an uncovered vertex which has inner degree $\geq n\big(1 - 1/(4K)\big)/2$.

R2. If x is mapped to the center of a star, y_i is unmapped and $\lambda(x) = \lambda(y_i)$ then y_i is mapped to an uncovered leaf of the star.

R3. If x is mapped to a leaf of a star such that the center of the star is uncovered, y_i is unmapped and $\lambda(x) = \lambda(y_i)$, then we map y_i onto the center of the star.

R4. We never cover the center of a star by a vertex x if at least one leaf of the star is already covered by a vertex y, for which $xy \notin E(T)$.

It is easy to satisfy R2 and R3. R4 can be satisfied in the course of the embedding of SP if $\delta(G) - n/2 \geq 2v(SP)$. For R1 observe that by Lemma 22 at most $12K^2\gamma n$ vertices have inner degree less than $n\big(1 - 1/(4K)\big)/2$, and so out of the at least $n/4 - v(SP) - 12K^2\gamma n > n/5$ choices one can always find an uncovered vertex for y_i with sufficiently large inner degree. In this way, every vertex of SP which is at a distance of at least four from every split vertex is mapped onto a vertex of G with large inner degree.

The distance four condition in the definition of SP is needed for the following reason. Assume x is a split vertex with, say, $\lambda(x) = A$, $xy \in E(SP)$ and $\lambda(y) = B$. When mapping y to B, it is possible that it is mapped to a leaf of a star. When we apply R3 and R2, we find that only a vertex which is of distance four from x can be mapped onto a vertex of large inner degree.

The problem we may face after finishing the embedding of SP is that some of the stars may have been damaged. That is, we may have covered some leaves of certain stars such that the center vertices are still uncovered. We solve this problem by finding new leaves for such center vertices. We accomplish this using the same method applied to find the stars initially. Say that there is an uncovered vertex $u \in B'$ such that some of the leaves of its star were covered. Assuming that $c_K \log n \geq v(SP) + 2K$, every vertex will have unmapped outer degree of at least $2K$ even after embedding SP. Hence, either we can find K vacant uncovered neighbors in B having inner degree at least $n/4$, or by averaging we will find an uncovered $v \in A$ such that v has at least K vacant uncovered neighbors, all having inner degree at least $n/4$.

In summary, SP is embedded such that every vertex which is at a distance of at least four from every split vertex is mapped onto a vertex with inner degree at least $n\big(1 - 1/(4K)\big)/2$. Moreover, if $x \in V(T) - V(SP)$, $y \in V(T)$, and $xy \in E(T)$ then $\lambda(x) = \lambda(y)$. That is, every subtree in the leftover is labelled to be mapped entirely to A or to B.

Embedding T – Step 1. First we state a simple but useful lemma.

Lemma 23. *Let $u, v \in B$. Then there is a $u - v$-path of length at most five entirely in B. The path can be found even if we discard at most $n/5$ vertices from B, as long as the stars of u and v remain intact. An analogous statement holds in case $u, v \in A$.*

Proof. We construct a $u - v$-path of length five when $u, v \in B'$. Let u_1 be an arbitrary leaf of the star centered at u and let v_1 be a leaf of the star centered at v. Since the inner degree of u_1 and v_1 each is at least $n/4$, even after deleting $n/5$ vertices both will have at least $n/20$ neighbors left in B. Out of these vertices in the neighborhood of u_1 we can find a vertex u_2 with inner degree at least $n/2 - 80\gamma n$ (see Lemma 22). Therefore, u_2 has a neighbor $v_2 \in N_G(v_1) \cap B$. The $u - v$-path of length five is: u, u_1, u_2, v_2, v_1, v.

If at least one of u, v belong to $B - B'$, then one can find an even shorter path, as is the case when both vertices belong to A. We leave the proof of these cases to the reader. ■

We find two subtrees of T, denoted by T_A and T_B such that $V(T_A) \subset A$ and $V(T_B) \subset B$, and $n/K^7 < v(T_A), v(T_B) \leq n/K^5$. We embed these subtrees into A or B by covering every vertex with inner degree at most $n\big(1 - 1/(3K)\big)/2$, thereby preparing G for the application of the general tree embedding lemma, Lemma 21.

Recall the decomposition of T from Lemma 17. After the first splitting, the smaller forest is assigned to A. It has size at least $n/3$, hence its largest subtree T_1 is of size at least $n/(3K)$. Then we split the leftover, and the smaller forest is assigned to B. The size of it is at least $n/6$, hence, its largest subtree T_2 has size at least $n/(6K)$. We find $T_A \subset T_1$ and $T_B \subset T_2$.

Finding T_A and T_B are very similar, we discuss finding T_A. Since $v(T_1) > n/(3K)$, T_A will be a proper subtree of T_1. Let $r \in V(T_1)$ be the unique vertex in $V(T_1)$ which has a neighbor in SP (if there were two neighbors, we would get a cycle). Beginning at r, form a maximal path $r = x_0, x_1, x_2, \ldots, x_k$ such that $T_1(x_i)$ is the largest subtree of T_1 rooted at a child of x_{i-1} and such that $v\big(T_1(x_i)\big) > n/K^5$ for every $i \leq k$. Choose z such that $T_1(z)$ is the largest subtree among the children of x_k. We let $T_A = T_1(z)$ and let z be the root of T_A. Clearly, $n/K^7 < v(T_A) \leq n/K^5$. In the same way we determine $T_B \subset T_2$ such that $n/K^7 < v(T_B) \leq n/K^5$. Observe, that $T_1 - T_A$ has only one vertex which is adjacent to T_A, and similarly for T_B. Since we put T_2 into B after two splittings, there may

be two edges from T_B to the rest of T, one connecting to $T_2 - T_B$ and one connecting to SP.

We discuss the algorithm for embedding T_A and T_B. This time we give the details for T_B, which is somewhat harder than the case of T_A.

We begin with putting the uncovered vertices of B in increasing order according to their inner degrees, writing $u_1, u_2, \ldots, u_\ell$, where $n/2 - v(SP) \leq \ell < n/2$. Thus we may refer to the first m vertices of B, by which we mean the set $\{u_1, u_2, \ldots, u_m\}$. The goal is to cover as many vertices from the beginning of this ordered list as possible.

We first map the root of T_B onto an uncovered vertex which has inner degree $\geq n\big(1 - 1/(4K)\big)/2$. Since T_B may have a vertex which is adjacent to SP, we are careful to map that vertex onto an uncovered vertex having inner degree $\geq n\big(1 - 1/(4K)\big)/2$.

A simple general rule will always be applied: when mapping x we map it on an uncovered vertex u such that $uI(y_i) \in E(G)$ for every mapped neighbor y_i. Apart from this obvious requirement we obey rules R2 and R3, as described during the embedding of SP.

Say that in the course of the embedding the most recently mapped vertex is x_i onto u_j for some $1 \leq i \leq j$, and that the first uncovered vertex of B according to the ordering is u_t, where $i \leq t$. Our next goal is to cover u_t. By Lemma 23 there is a path of length at most five between u_j and u_t, the length of which is denoted by ℓ.

Consider the induced subtree containing unmapped vertices of $T_B(x_i)$, denoted $T'_B(x_i)$, having depth ℓ and being comprised of every vertex below x_i which is at distance at most ℓ from x_i. This subtree has at most K^5 vertices, which we embed greedily, but in such a way that one of its leaves will cover u_t. By Lemma 23 this is possible to achieve. Then we pick the next vertex in the ordering of $V(B)$, covering which may again embed at most K^5 vertices from T_B. That is, by mapping at most K^5 vertices of the tree, we can cover at least one uncovered vertex of B'. Hence, after mapping at most $K^5|B'| \leq K^5 16\gamma n$ vertices of T_B we will have no uncovered vertex left in B'.

After covering every vertex of B' we will try to cover as many "small" inner degree vertices as possible. Again, we apply Lemma 23 and find that by mapping at most K^5 vertices of T_B we can cover any uncovered vertex we want. An easy calculation shows that we can cover the first $n/(2K^{12})$ vertices of B in this way. By Lemma 22 every uncovered vertex left in B will have inner degree at least $n/2 - 4K^{-8}n > n\big(1 - 1/(4K)\big)/2$.

Observe that when mapping the vertices of T_A we get at least the same bounds for the inner degrees of uncovered vertices of A.

Remark 6. After this step every uncovered vertex has inner degree at least $n\big(1-1/(4K)\big)/2$. We covered at most n/K^5 vertices of A and of B, hence the minimum degree inside these sets is at least $n\big(1-1/(4K)\big)/2 - n/K^5$.

Embedding T – Step 2. It is now a simple task to finish the embedding of T by the help of Lemma 21. Denote the forest induced by unmapped vertices which have to be mapped into A by $\widehat{T}$, and let $\widetilde{T}$ denote the forest of unmapped vertices which have to be mapped into B.

Take an arbitrary proper two coloring of $\widehat{T}$, letting Q and R denote the color classes, and set $q = |Q|$ and $r = |R|$. Divide the uncovered vertices of A randomly into two sets, Q_A and R_A, where $|Q_A| = q$ and $|R_A| = r$.

Using Azuma's inequality and Remark 6 every $u \in Q_A$ will be adjacent to at least $r\big(1-1/(3K)\big)$ vertices in R_A, and every $v \in R_A$ will be adjacent to at least $q\big(1-1/(3K)\big)$ vertices in Q_A. There are at most $v(SP)+1$ vertices which are a priori restricted to be mapped onto given vertices of A – the vertices of SP and the unique vertex in T_1 adjacent to T_A. Since the requirements of Lemma 21 are satisfied, we conclude that $\widehat{T}$ can be embedded. $\widetilde{T}$ is embedded similarly. Hence, we conclude that in the first extremal case $T \subset G$. ■

Remark 7. Note that $c_K = 4K^4$ is sufficiently large in the first extremal case.

3.2. The Second Extremal Case

In this case G is close to $K_{n/2,n/2}$. This case is in fact very similar to the previous one, so the emphasis will be on the differences. We discuss the switching procedure in detail, and give an outline of the rest of the embedding method.

We assume that $W(G)$ is partitioned into two vertex classes of size $n/2$, A and B, such that $e(A) + e(B) \leq \gamma n^2$. Observe that by the minimum degree condition, every vertex will have at least $c_K \log n$ neighbors in both vertex classes.

We need a very similar statement to Lemma 22:

Lemma 24. *Let $0 < \eta < 1/2$. Then the number of vertices of A having less than $n/2 - \eta n$ neighbors in B is at most $\frac{2\gamma n}{\eta}$. The analogous statement is true for B.* ■

Preparations for the embedding – Step 1: Decomposition of T. We start by finding a decomposition of the tree T by Lemma 18 into a set of subtrees $T_1, T_2, \ldots, T_s$ with the following properties: (1) $s = O(\log n)$, (2) there is a proper red-blue coloring of these subtrees such that overall we color $n/2$ vertices red and $n/2$ vertices blue and (3) there are at most $O(\log n)$ edges of T inside the color classes, each of which is incident to a split vertex. As before, if $x \in V(T)$ then $\lambda(x) = A$ or B indicates into which set we map x. It is easy to see that the largest subtree of this decomposition, T_1 has at least $n/(3K)$ vertices.

Given the decomposition of T, we define the forest SP as in the previous extremal case as the induced sub-forest of T containing every vertex being at a distance at most four from a split vertex. By the definition of SP, there are no two components of SP with an edge connecting them.

Preparations for the embedding – Step 2: Switching and finding stars. We repeatedly perform a switching procedure whenever it is possible. If we identify a vertex $v \in A$ and a vertex $u \in B$ such that $e(A, B) < e(A - v + u, B - u + v)$, then we let $A = A - v + u$ and $B = B - u + v$. Since the number of edges in between A and B increases at every step of the switching, this is a finite process. Note also that when we are finished with the switching, one cannot find a pair of vertices $v \in A$, $u \in B$ such that $\left|N(v) \cap B\right| \leq n/4$ and $\left|N(u) \cap A\right| \leq n/4$. Therefore, in one of the vertex classes, say A, we find only vertices which have at least $n/4$ neighbors in the other vertex class, B. We will use the notion of the *inner* and *outer* degree, as defined in the previous extremal case.

We pay special attention to those vertices of B which have very small outer degree. Observe that we may have at most $8\gamma n$ vertices in B each having outer degree less than $n/4$. Denote their set by B'.

As before, we find $|B'|$ vertex disjoint stars on $K + 1$ vertices. The leaves of these stars come from A, and the center vertices are either from A or from B. We determine such stars as follows. Let A_v be the set of vacant vertices, and set $A_v = A$ and $A' = \emptyset$ at the outset. One-by-one we look for K vacant neighbors for every $u \in B'$. We have the following cases.

Case 1: If u is adjacent to $v_1, v_2, \ldots, v_K \in A_v$, then the new star will have center u and leaves $v_1, v_2, \ldots, v_K$. We let $A_v = A_v - \{v_1, v_2, \ldots, v_K\}$, and $A' = A' + \{v_1, v_2, \ldots, v_K\}$.

Case 2: If u does not have K vacant neighbors, but there is a star with center $v \in A_v$ and leaves $v_1, v_2, \ldots, v_K \in A_v$, then we switch u with v. The new star will have center v and leaves $v_1, v_2, \ldots, v_K$. We let $A_v = A_v + u - \{v_1, v_2, \ldots, v_K\}$ and $A' = A' + \{v_1, v_2, \ldots, v_K\}$.

Lemma 25. *Assume that there exists a $u \in B'$ such that it does not have K vacant neighbors. Then A_v contains a star on $K+1$ vertices.*

Proof. Suppose we have found t stars. Clearly, $t > \log_n$ by the minimum degree condition on G. Let us assume to the contrary that A_v does not contain a star on $K+1$ vertices.

If u does not have K neighbors in A_v then u has at most $Kt+t+K-1$ neighbors in A after the first switching procedure is concluded. At most Kt of these neighbors are from A', at most $K-1$ are from A_v and we apply the switching of Case 2 at most t times. Since $t > K$, we have that $Kt+t+K-1 \leq (K+2)t$. The following claim is easy to prove:

Claim 26. *If at the end of the first switching procedure B has a vertex with outer degree d, then no vertex of A has inner degree larger than $2d$.* ■

This implies that the vertices of A' have inner degree at most $2(K+2)t$. Assuming A_v has no star on $K+1$ vertices, we have the following lower bound on the number of edges in between A and B:

$$e(A,B) \geq \frac{n^2}{4} - 2K(K+2)t^2 - (K-1)\frac{n}{2}.$$

We also have the following upper bound on $e(A,B)$ by inspecting the outer degrees of the vertices of B:

$$e(A,B) \leq t\frac{n}{4} + \left(\frac{n}{2} - t\right)\frac{n}{2.}$$

From these we get the inequality

$$0 \leq 2K(K+2)t^2 - \frac{n}{4}t + (K-1)\frac{n}{2}.$$

This inequality is not satisfied for $t \in [2K, 8\gamma n]$. ■

Hence, if we fail to apply the method of Case 1 then Case 2 can always be applied, and the stars can be determined.

Preparations for the embedding – Step 3: Embedding SP. Embedding SP is very similar to the previous case. Our goal is to map x onto an uncovered vertex $u \in \lambda(x)$ such that if $xy \in E(T)$ and we have already mapped y then $uI(y) \in E(G)$. We begin mapping a component of SP by its root. As in the previous extremal case, we have four rules concerning the embedding of SP.

R1′. If x is mapped to a vertex with outer degree at least $n/4$, y_i is unmapped and $\lambda(x) \neq \lambda(y_i)$, then we map y_i onto an uncovered vertex which has outer degree $\geq n\big(1 - 1/(4K)\big)/2$.

R2′. If x is mapped to the center of a star, y_i is unmapped and $\lambda(x) \neq \lambda(y_i)$ then we map y_i to an uncovered leaf of the star.

R3′. If x is mapped to a leaf of a star such that the center of the star is uncovered, y_i is unmapped and $\lambda(x) \neq \lambda(y_i)$, then we map y_i onto the center of the star.

R4′. We never cover a center of a star by a vertex x if at least one leaf of the star is already covered by a vertex y for which $xy \notin E(T)$.

As before, it is easy to satisfy the above rules, and as before we may face the problem that after finishing the embedding of SP, some of the stars may have been damaged. That is, we may have covered some leaves of certain stars without covering their centers. We solve this problem by finding new leaves for such center vertices. Suppose that there is an uncovered vertex $u \in B'$ such that some of the leaves of its star were covered. Assuming that $c_K \log n \geq v(SP) + 2K$, every vertex will have an uncovered outer degree of at least $2K$ even after embedding SP. Hence, either we can find K vacant uncovered neighbors in B having outer degree at least $n/4$, or by averaging we will find an uncovered $v \in A$ such that v has at least K vacant uncovered neighbors, all having outer degree at least $n/4$, by Lemma 25. In this way we find the missing leaves for the damaged stars.

In summary, after embedding SP every vertex which is of distance at least four from every split vertex is mapped onto a vertex which has outer degree at least $n\big(1-1/(4K)\big)/2$. Moreover, if $x \in V(T)-V(SP)$, $y \in V(T)$, and $xy \in E(T)$, then $\lambda(x) \neq \lambda(y)$. That is, λ represents a proper 2-coloring for every subtree in the leftover.

Embedding T – Step 1. First we state a simple but useful lemma whose proof is very similar to the proof of Lemma 23.

Lemma 27. *Let $u, v \in B$. Then there is a $u - v$-path of length at most six, with vertices alternately belonging to B and A. The path can be found even after discarding at most $n/5$ vertices, as long as the stars of u and v remain intact. The analogous statement holds in case $u, v \in A$.* ■

As in the previous extremal case, we find a subtree of T, denoted by T' such that $n/K^7 < v(T') \leq n/K^5$. As in the previous extremal case, when embedding this subtree we cover every vertex with outer degree at most $n\big(1 - 1/(3K)\big)/2$, thereby preparing G for the application of the general tree embedding lemma, Lemma 21. Recall that the largest subtree in the decomposition, T_1, has at least $n/(3K)$ vertices.

Again, we order the uncovered vertices of $W(G)$ according to increasing outer degree and continue the embedding of T by that of T'. In order to cover the low outer degree vertices we will repeatedly apply Lemma 27, achieving at the end that we cover the first $n/(2K^{13})$ vertices in the list. We omit the details as they are very similar to the one discussed in the first extremal case. By Lemma 24 every uncovered vertex left in $W(G)$ has outer degree at least $n/2 - 4K^{-7}n > n\big(1 - 1/(4K)\big)/2$.

Remark 8. After this step every uncovered vertex has outer degree at least $n\big(1 - 1/(4K)\big)/2$. We covered at most n/K^5 vertices of $W(G)$, and so the minimum degree in the leftover is at least $n\big(1 - 1/(4K)\big)/2 - n/K^5$.

Embedding T – Step 2. In order to finish the embedding of T we observe that the conditions of Lemma 21 are satisfied for the leftover forest $\widehat{T}$, and so $\widehat{T}$ can be embedded. Hence, we conclude the second extremal case. ■

Remark 9. Notice, that $c_K = 4K^4$ is sufficiently large in this extremal case as well.

4. Lower Bound for the Minimum Degree in G

In this section we show that $\delta(G) - n/2 = \Omega(\log n)$ for both extremal cases. First we need a definition.

Definition 2. A balanced cut of a graph is a partitioning of its vertex sets into two subsets, such that their sizes differ by at most one. The size of a cut is the number of edges going between the two partition sets.

As we will see, the minimum size of a balanced cut of T provides a lower bound when embedding it in the first extremal case. while, the maximum size cut provides a lower bound for the second extremal case. In both cases T is a complete ternary tree on n vertices, and the root r of T is the only vertex having degree three, all other vertices have degree 4 or 1. We define the height $h(x)$ of x as the number of edges on an $x - y$ path, where y is the closest leaf to x.

4.1. The First Extremal Case

Lemma 28. *Let T be a complete ternary tree of height h. Then the size of any balanced cut of T is at least $\lfloor h/2 \rfloor$.*

Proof. Let A and B be the partition sets of a minimum balanced cut of T, we assume that $|A| \geq |B|$. Let $\chi : V(T) \to \{-1, 1\}$ be a coloring function. We initialize the coloring $\chi(x) = 1$ for every x in $V(T)$, then apply a *color-swapping process* such that at the end of the process $A = \chi^{-1}(1)$ and $B = \chi^{-1}(-1)$.

> The *Color-swapping Process:*
>
> Initially the color of every vertex is 1. Traverse T by a depth-first search. If there is a vertex x such that x and its ancestor are not in the same partition set then multiply the colors of the vertices of $T(x)$ by -1. We call this operation the *color-swap* of $T(x)$.

At the end of the process we get the desired coloring χ such that $A = \chi^{-1}(1)$ and $B = \chi^{-1}(-1)$. It is clear that this process will always yield a coloring which corresponds to the cut, and that the number of color-swaps performed is the size of the cut. For any coloring χ, label every vertex x by the signed difference of the size of the color classes of $T(x)$. More precisely, for every x let $L(x) = \sum_{y \in T(x)} \chi(y)$. In the procedure described, the labels are updated whenever a color-swap operation is performed.

When we swap the colors of $T(x)$, the label of a vertex y on the $r - x$ path changes by adding $-2L(x)$ and the labels of the vertices of $T(x)$ are

multiplied by -1. It is useful to represent the labels in a modified ternary numeral system in which the digits are -1, 0, 1. Such a representation will be subscripted by a "3". For example, $101(-1)_3 = 3^3 + 3^1 - 1 = 29$. Every integer can be represented this way and this representation is unique. Let $\ell_k = 11\ldots1_3$ where the subscript k refers to the number of 1s. The following facts hold:

We introduce the notation $\|a\|$ for an integer a, defined to be the number of non-zero digits in the modified ternary representation of a. It is easy to see (by induction on the number of digits of a) that if $|b| \leq |a|$ then $\|a \pm b\| \leq \|a\| + \|b\|$. We refer to $\|a\|$ as the *norm* of a.

1. Initially the label of any $x \in T$ is $L(x) = \ell_k$ where $k = h(x) + 1$. Hence, initially $\|L(x)\| = h(x) + 1$.

2. At the end of the process $\|L(r)\| \leq 1$.

3. After a color-swap of $T(x)$ the labels of the vertices on the $r - x$ path increase or decrease by $100\ldots0(-1)_3$. If $\chi(x) = 1$, then the labels decrease, otherwise increase. Here the number of 0s is $h(x)$.

Initially $\|L(r)\| = h(r) + 1$, and at every color-swap $L(r)$ increases or decreases by a number which has norm two. If the size of the minimum balanced cut is m, then there are m color-swaps. Since one color-swap can decrease $\|L(r)\|$ by at most two, at least $\lfloor h/2 \rfloor$ color-swaps are needed in order for $\|L(r)\| \leq 1$ at the end. The lemma follows. ■

We define the extremal graph $G = (W, E_G)$. Set $h = \log_3 n$, $n_1 = n/2 - h/16 + 1$ and $n_2 = h/16 - 1$. The vertex set of G is decomposed into four disjoint subsets, $W = X_1 \cup X_2 \cup Y_1 \cup Y_2$, where X_1 and X_2 are both n_1-cliques. Let Y_1 and Y_2 each be sets on n_2 vertices and let $Y = Y_1 \cup Y_2$ be an independent set. Finally, we form the complete bipartite graphs between (X_1, Y) and (X_2, Y). It is easy to see that $\delta(G) = n_1 - 1 + 2n_2 = n/2 + h/16 - 2$.

Assume there is an embedding $I : V \to W$ of the complete ternary tree T into G. In order to apply Lemma 28, let $A = I^{-1}(X_1 \cup Y_1)$ and $B = I^{-1}(X_2 \cup Y_2)$. We have $|A| = |B| = n/2$, and therefore $e_T(A, B) \geq h/2$. Since $e_G(X_1, X_2) = 0$, one endpoint of every edge going in between A and B is mapped to a vertex in Y. The maximum degree of T is 4, hence there are at most $4|Y|$ edges of T with one endpoint mapped onto some vertex in Y. Since $4|Y| < h/2$ we have arrived at a contradiction: $T \not\subset G$.

4.2. The Second Extremal Case

Lemma 29. *Let T be a complete ternary tree with height h. The size of any balanced cut of T is at most $e(T) - \lfloor h/2 \rfloor$.*

Proof. Let A and B, $|A| > |B|$, be the vertex sets of a balanced maximum cut of T. We show that the number of those edges which have both endpoints belonging to the same set is at least $\lfloor h/2 \rfloor$ in a manner similar to the previous case. We apply a color-swapping process similarly to the previous case, but with two modifications. First, the coloring function $\chi : V(T) \to \{-2, 2\}$ is initialized as a proper 2-coloring of T with $\chi(r) = 2$. Second, we multiply the vertices of $T(x)$ by -1 when x and its ancestor are in the *same* partition set. After applying the Color-swapping Process, we have the following:

1. The labels of the vertices of T depend on their height and the height of the tree:

   (a) If $h(r)$ and $h(x)$ are even, then $L(x) = 1(-1)\dots(-1)_3$.

   (b) If $h(r)$ is even and $h(x)$ is odd, then $L(x) = 11\dots1_3$.

   (c) If $h(r)$ is odd and $h(x)$ is even, then $L(x) = (-1)11\dots1_3$.

   (d) If $h(r)$ and $h(x)$ are odd, then $L(x) = (-1)(-1)\dots(-1)_3$.

   In each case the number of digits of $L(x)$ depends on the parity of $h(x)$. If $h(x)$ is even, then the number of digits is $h(x) + 2$, otherwise it is $h(x) + 1$.

2. At the end of the process $L(r) = 0$ or $1(-1)_3$, depending on the parity of h.

3. After the color-swap of $T(x)$ the labels of the vertices on the $r - x$ path increase or decrease. If $h(x)$ is even, then the absolute value of the change is $10\dots01_3$, where the number of zeros is $h(x)$. If $h(x)$ is odd, then the labels change by $10\dots0(-1)_3$, here the number of zeros is $h(x)$.

The above facts can be proved easily by induction. Application of the same reasoning as before implies the statement of the lemma. ∎

This time the extremal graph $G = (W, E_G)$ is defined as follows. First, set $h = \log_3 n$, $n_1 = n/2 - h/16 + 1$ and $n_2 = h/16 - 1$. The vertex set of G is decomposed into four disjoint subsets: $W = X_1 \cup X_2 \cup Y_1 \cup Y_2$, where X_1 and X_2 are two independent sets of size n_1, and $G|_{X_1 \cup X_2}$ is isomorphic to K_{n_1,n_1}. Let Y_1 and Y_2 be two sets on n_2 vertices, and let $Y = Y_1 \cup Y_2$ be an independent set. Finally, let $G|_{X_1 \cup Y}$ and $G|_{X_2 \cup Y}$ be isomorphic to $K_{n_1,2n_2}$. It is easy to see that $\delta(G) = n_1 + 2n_2 = n/2 + h/16 - 1$.

As before, let $I : V \to W$ denote the embedding of the complete ternary tree T into G. We apply Lemma 29 to the cut defined by $A = I^{-1}(X_1 \cup Y_1)$ and $B = I^{-1}(X_2 \cup Y_2)$. Obviously, $|A| = |B| = n/2$, and therefore $e_T(A, A) + e_T(B, B) \geq h/2$. Assume, that $e_T(A, A) \geq e_T(B, B)$. By construction, one vertex of every edge of T mapped into A is mapped to a vertex of Y_1. The maximum degree of T is 4, and so there are at most $4|Y_1|$ edges of T with one endpoint mapped onto some vertex in Y_1. Since $4|Y_1| < h/4$ we have arrived at a contradiction: $T \not\subset G$.

References

[1] N. Alon and J. Spencer, *The Probabilistic Method,* John Wiley & Sons, Hoboken, New Jersey (2008).

[2] B. Bollobás, *Extremal Graph Theory,* Academic Press, London (1978).

[3] J. Komlós, G. N. Sárközy and E. Szemerédi, Proof of a packing conjecture of Bollobás, *Combinatorics, Probability and Computing,* **4** (1995), 241–255.

[4] J. Komlós, G. N. Sárközy and E. Szemerédi, Spanning trees in dense graphs, *Combinatorics, Probability and Computing,* **10** (2001), 397–416.

Béla Csaba

Department of Mathematics
Western Kentucky University

e-mail: bela.csaba@wku.edu

Ian Levitt

Department of Mathematics
Rutgers University

e-mail: ianl@math.rutgers.edu

Judit Nagy-György

Bolyai Institute
University of Szeged

e-mail:

nagy-gyorgy@math.u-szeged.hu

Endre Szemerédi

Department of Computer Sciences
Rutgers University
and
Rényi Institute
Budapest
and
IAS, Princeton

e-mail: szemered@cs.rutgers.edu

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 139–149.

BETTI NUMBERS ARE TESTABLE*

GÁBOR ELEK

We prove that the Betti numbers of simplicial complexes of bounded vertex degrees are testable in constant time.

1. INTRODUCTION

Property testing in bounded degree graphs was introduced in the paper of Goldreich and Ron [4]. In this paper we study property testing for bounded degree simplicial complexes in higher dimensions. Let $d \geq 2$ be a natural number and consider finite simplicial complexes where each vertex (zero dimensional simplex) is contained in at most d edges (1-dimensional simplex). Of course, such a complex can be at most d-dimensional. What does it mean to test the p-th Betti number of such a simplicial complex? First fix a positive real number $\varepsilon > 0$. A tester takes a simplicial complex K as an input and pick $C(\varepsilon)$ random vertices. Then it looks at the $C(\varepsilon)$-neighborhoods of the chosen vertices. Based on this information the tester gives us a guess $\hat{b}^p(K)$ for the p-th Betti number $b^p(K)$ of the simplicial complex such a way that:

$$\text{Prob}\left(\frac{\left|\hat{b}^p(K) - b^p(K)\right|}{\left|V(K)\right|} > \varepsilon\right) < \varepsilon,$$

where $V(K)$ is the set of vertices in K. In other words, we can estimate the p-th Betti number very effectively with high probability knowing only

*Research sponsored by OTKA Grant K 69062 and 49841.

a small (random) part of the simplicial complex. The goal of this paper is to show the existence of such a tester for any $\varepsilon > 0$. That is to prove the following theorem:

Theorem 1. *Betti-numbers are testable for bounded degree simplicial complexes.*

For graphs the 0-th Betti number is just the number of components and the first Betti number can be computed via the 0-th Betti number and the Euler-characteristic, hence it is not hard to see that such tester exists. For connected surfaces one can also calculate the first Betti number using just the number of vertices, edges and triangles. However in higher dimensions there is no such formula even for triangulated manifolds. Note that this paper was not solely motivated by the paper of Goldreich and Ron, but also by the solution of the Kazhdan–Gromov Conjecture by Wolfgang Lück [5]. The workhorse lemma of our paper is basically extracted from his paper using a slightly different language. It is very important to note that our proof works only for Betti numbers of real coefficients and we do not claim anything for the Betti numbers of mod-p coefficients.

2. The Convergence of Simplicial Complexes

Let Σ^d be the set of finite simplicial complexes K of vertex degree bound d that is any 0-dimensional simplex is contained in at most d 1-dimensional simplices. We denote by K_i the set of i-simplices in K and by G_K the 1-skeleton of K, that is $V(G_K) = K_0$, $E(G_K) = K_1$. A rooted r-ball of degree bound d is a simplex $L \in \Sigma^d$ with a distinguished vertex x such that for any $y \in V(G_L)$, $d(x,y) \leq r$, where $d(x,y)$ is the shortest path distance of x and y in the graph G_L. We denote by $Z^{r,d}$ the rooted isomorphism classes of rooted r-balls. If $K \in \Sigma^d$ and $p \in V(K)$ then let $G_r(p)$ be the rooted r-ball in the 1-skeleton G_K and $B_r(p)$ is the set of simplices σ such that all vertices of σ are in $G_r(p)$. Then $B_r(p)$ is a rooted r-ball of vertex degree bound d.

For $\alpha \in Z^{r,d}$ we denote by $T(K,\alpha)$ the set of vertices p such that $B_r(p) \cong \alpha$. We set

$$p_K(\alpha) := \frac{|T(K,\alpha)|}{|K_0|}.$$

We say that $\{K^n\}_{n=1}^{\infty} \subset \Sigma^d$ is convergent (see [1] for the graph case) if

- $|K_0^n| \to \infty$
- $\lim_{n\to\infty} p_{K_n}(\alpha)$ exists for any $r \geq 1$ and $\alpha \in Z^{r,d}$.

It is easy to see that any sequence $\{L^n\}_{n=1}^{\infty} \subset \Sigma^d$ such that $|L_0^n| \to \infty$ contains a convergent subsequence.

Let $\alpha_1, \alpha_2, \alpha_3, \ldots$ be an enumeration of all the r-balls, $r \geq 1$. Then we have the pseudo-metric

$$d_s(K, L) := \sum_{i=1}^{\infty} \frac{1}{2^i} \left| p_K(\alpha_i) - p_L(\alpha_i) \right|.$$

Clearly, if $\{K^n\}_{n=1}^{\infty}$ be an increasing sequence of simplicial complexes they are Cauchy if and only if they are convergent. By an oriented d-complex Q we mean an element of Σ^d equipped with a fixed orientation for each of its simplex. Note that we do not assume that the orientations are compatible in any sense. We denote by $\hat{\Sigma}^d$ the set of all finite oriented d-complexes. We also define oriented r-balls, the set $\hat{Z}^{r,d}$ of all oriented r-ball isomorphism classes and the probabilities $p_Q(\beta)$ accordingly. Naturally we can define the convergence of oriented d-complexes as well.

Proposition 2.1. *Let $\{K^n\}_{n=1}^{\infty} \subset \Sigma^d$ be a convergent sequence of d-complexes. Then one has an oriented copy Q^n for each K^n such that $\{Q^n\}_{n=1}^{\infty} \subset \hat{\Sigma}^d$ is convergent as well.*

Proof. Consider i.i.d. random variables $\theta(x)$ distributed uniformly on $[0, 1]$. Let $(a_0, a_1, \ldots, a_i) \in K_i^n$ be a positive orientation if

$$\theta(a_0) < \theta(a_1) < \cdots < \theta(a_i).$$

Then the resulting sequence $\{Q^n\}_{n=1}^{\infty} \subset \hat{\Sigma}^d$ is convergent with probability 1. For details of the simple argument see ([3], Proposition 2.2). ■

Finally, we need a technical definition that we use in the subsequent sections. Let us consider the set Q_i of i-simplices in an oriented d-complex Q. If $\sigma, \tau \in Q_i$ we say that σ and τ are adjacent if they have at least one joint vertex. This way we can define the shortest path distance $d_i(\sigma, \tau)$ (where $d_i(\sigma, \tau) = \infty$ is possible). The ball $B_r^i(\sigma)$ is the set of i-simplices τ such that $d_i(\sigma, \tau) \leq r$. As above, we can define the classes $\hat{Z}_i^{r,d}$ and the sampling probabilities $p_{Q_i}(\beta)$. Note that if $\alpha, \beta \in \hat{Z}_i^{r,d}$ we say that α is isomorphic to β if they are isomorphic as simplicial complexes not only as metric spaces.

That is $p_{Q_i}(\beta)$ defined the following way. Let $\beta \in \hat{Z}_i^{r,d}$. Denote by $T(Q_i, \beta)$ the number of i-simplices τ such that the simplicial complex $B_r^i(\sigma)$ is isomorphic to β, where the isomorphism preserves the root-simplex. Then

$$p_{Q_i}(\beta) := \frac{|T(Q_i, \beta)|}{|Q_i|}.$$

Clearly, if $\{Q^n\}_{n=1}^{\infty} \subset \hat{\Sigma}^d$ is a convergent sequence then for any $\beta \in \hat{Z}_i^{r,d}$, $\lim_{n\to\infty} p_{Q_i^n}(\beta)$ exists.

3. Betti Numbers and Combinatorial Laplacians

Let $Q \in \hat{\Sigma}^d$ be an oriented simplicial complex. Let $C^i(Q)$ denote the euclidean space of real functions on the 1-simplices of Q. Let us consider the cochain-complex

$$C^0(Q) \xrightarrow{d_0} C^1(Q) \xrightarrow{d_1} \dots$$

Recall that if $f \in C^q(Q)$ then

$$df(a_0, a_1, \dots, a_{q+1}) = f(a_1, a_2, \dots, a_q) - f(a_0, a_2, \dots, a_{q+1}) + \dots$$
$$+ (-1)^{q+1} f(a_0, a_1, \dots, a_q).$$

Then $b^i(Q) = \dim \operatorname{Ker} d_i - \dim \operatorname{Im} d_{i-1}$ are the Betti numbers of Q. Note that they do not depend on the choice of the orientation of Q only the underlying simplicial complex.

The combinatorial Laplacians (see e.g. [2]) $\Delta_Q^i : C^i(Q) \to C^i(Q)$ are defined as

$$\Delta_Q^i := d_{i-1} d_{i-1}^* + d_i^* d_i.$$

The operators Δ_Q^i are positive and self-adjoint. Also,

$$\dim \operatorname{Ker} \Delta_Q^i = b^i(Q).$$

Let us remark that by Lemma 2.5 of [2] we have the following information on the combinatorial Laplacians:

- $\Delta_Q^i(\sigma, \tau) \neq 0$ only if $\sigma = \tau$ or σ and τ are adjacent.
- $\Delta_Q^i(\sigma, \tau)$ is always an integer.
- $|\Delta_Q^i(\sigma, \tau)| \leq d + 1$.

4. Weak Convergence of Probability Measures

First recall the notion of weak convergence of probability measures. Let $\{\mu_n\}_{n=1}^{\infty}$ be probability measures on the interval $[0, K]$. Then $\{\mu_n\}_{n=1}^{\infty}$ weakly converges to μ if for any continuous function $f \in C[0, K]$

$$\int_0^K f\, d\mu_n \to \int_0^K f\, d\mu.$$

For an example, let $\mu_n(\frac{1}{n}) = 1$, then the measures μ_n converge to the measure concentrated at the zero. Note that in this case $\lim_{n\to\infty} \mu_n(0) \neq \mu(0)$.

Now let

$$c(\mu_n) = \lim_{\varepsilon\to 0} \int_{\varepsilon}^{K} \log \lambda\, d\mu_n.$$

The following theorem can be extracted from [5], nevertheless we provide a proof using only the language of real analysis, avoiding any reference to operators.

Theorem 2. *Suppose that* $\{\mu_n\}_{n=1}^{\infty}$ *weakly converges to* μ *and for any* $n \geq 1$, $c(\mu_n) \geq 0$. *Then* $\lim_{n\to\infty} \mu_n(0) = \mu(0)$.

Proof. First we need some notations. For a monotone function f,

$$f^+(\lambda) = \inf_{\varepsilon\to 0} f(\lambda + \varepsilon)$$

For the measures μ_n let σ_n be their distribution function that is

$$\sigma_n(\lambda) = \mu_n\big([0, \lambda]\big).$$

Also, let $t(\lambda) = \mu\big([0, \lambda]\big)$. Note that $\sigma_n^+ = \sigma_n$. Let

$$\overline{\sigma}(\lambda) := \limsup_{n\to\infty} \sigma_n(\lambda)$$

and

$$\underline{\sigma}(\lambda) := \liminf_{n\to\infty} \sigma_n(\lambda).$$

The following lemma trivially follows from the definitions.

Lemma 4.1. *Let f be a continuous function such that*

$$\chi_{[0,\lambda]}(x) \leq f(x) \leq \chi_{[0,\lambda+\frac{1}{k}]}(x) + \frac{1}{k}$$

for any $0 \leq x \leq K$, then

$$\sigma_n(\lambda) \leq \int_0^K f(\lambda)\, d\mu_n \leq \sigma_n(\lambda + \frac{1}{k}) + \frac{1}{k}.$$

$$t(\lambda) \leq \int_0^K f(\lambda)\, d\mu \leq t(\lambda + \frac{1}{k}) + \frac{1}{k}.$$

Proposition 4.1. $\overline{\sigma}(\lambda) \leq t(\lambda) = \underline{\sigma}^+(\lambda) = \overline{\sigma}^+(\lambda)$.

Proof. By this lemma,

$$\overline{\sigma}(\lambda) \leq t(\lambda) \leq \overline{\sigma}\left(\lambda + \frac{1}{k}\right) + \frac{1}{k}.$$

Hence $\overline{\sigma}(\lambda) \leq t(\lambda) = \underline{\sigma}^+(\lambda) \leq \overline{\sigma}^+(\lambda)$. Since $t(\lambda)$ is monotone, we have that $\overline{\sigma}(\lambda+\varepsilon) \leq t(\lambda+\varepsilon)$ and $\overline{\sigma}^+(\lambda) \leq t^+(\lambda) = t(\lambda)$. Thus our proposition follows. ■

The following elementary analysis lemma is proved in [6].

Lemma 4.2. *Let f be a continuously differentiable function on the positive reals and μ be a probability measure on the $[0, K]$ interval. Suppose that F is the distribution function of μ, that is $\mu[0, \lambda] = F(\lambda)$. Then for any $0 < \varepsilon \leq K$:*

$$\int_\varepsilon^K f(\lambda)\, d\mu = -\int_\varepsilon^K f'(\lambda)F(\lambda)\, d\lambda + f(K)F(K) - f(\varepsilon)F(\varepsilon).$$

Assume that $K \geq 1$, then by the previous lemma

$$c(\mu_n) = \log K - \log(\varepsilon)\sigma_n(\varepsilon) - \int_\varepsilon^K \sigma_n(\lambda)\frac{1}{\lambda}\, d\lambda.$$

Since

$$\int_\varepsilon^K \frac{\sigma_n(0)}{\lambda} = \big(\log K - \log(\varepsilon)\big)\,\sigma_n(0),$$

we have that

$$c(\mu_n) = \log K\big(1 - \sigma_n(0)\big) - \int_0^K \frac{\sigma_n(\lambda) - \sigma_n(0)}{\lambda}\, d\lambda.$$

That is

$$\int_0^K \frac{\sigma_n(\lambda) - \sigma_n(0)}{\lambda} \leq \log K. \tag{1}$$

Observe that

$$\int_\varepsilon^K \frac{\underline{\sigma}(\lambda) - \overline{\sigma}(0)}{\lambda}\, d\lambda \leq \int_\varepsilon^K \frac{\liminf_{n\to\infty}\big(\sigma_n(\lambda) - \sigma_n(0)\big)}{\lambda}\, d\lambda.$$

By Fatou's Lemma,

$$\int_\varepsilon^K \frac{\liminf_{n\to\infty}\big(\sigma_n(\lambda) - \sigma_n(0)\big)}{\lambda}\, d\lambda \leq \liminf_{n\to\infty} \int_\varepsilon^K \frac{\sigma_n(\lambda) - \sigma_n(0)}{\lambda}\, d\lambda. \tag{2}$$

Since the right hand side of (2) is less than $\log K$, we obtain the following inequality:

$$\int_0^K \frac{\underline{\sigma}(\lambda) - t(0)}{\lambda}\, d\lambda \leq \log K.$$

Therefore, $\lim_{\lambda\to 0} \underline{\sigma}(\lambda) = \overline{\sigma}(0)$. That is by Proposition 4.1

$$\overline{\sigma}(0) = t(0). \tag{3}$$

Since one can apply (3) for any subsequence of $\{\sigma_n\}_{n=1}^\infty$ we obtain that

$$\lim_{n\to\infty} \sigma_n(0) = t(0). \qquad \blacksquare$$

5. Spectral Convergence

The goal of this section is to prove the main technical proposition of our paper. Note this is based again on the ideas in [5]. Let $P : \mathbb{R}^n \to \mathbb{R}^n$ be a positive, self-adjoint operator and μ_P be its normalized spectral measure that is

$$\mu_P(\lambda) := \frac{\text{the multiplicity of } \lambda \text{ as an eigenvalue}}{n}.$$

Note that if $\|P\| \leq K$ then μ_P is concentrated on the interval $[0, K]$.

Proposition 5.1. *Let $\{Q^n\}_{n=1}^{\infty} \subset \hat{\Sigma}^d$ be a convergent sequence of oriented simplicial complexes. Then there exists $K > 0$ such that*

a) *For any $i \geq 1$ and $n \geq 1$, $\|\Delta^i_{Q^n}\| \leq K$.*

b) *The normalized spectral measures of $\{\Delta^i_{Q^n}\}_{n=1}^{\infty}$, $\{\mu^i_n\}_{n=1}^{\infty}$ weakly converge.*

c) *$c(\mu^i_n) \geq 0$ for any $i, n \geq 1$.*

Proof. To show (a) it is enough to prove the following lemma.

Lemma 5.1. *Let $L, M > 0$ be positive integers. Then if A is a $n \times n$-matrix of real coefficients (that is a linear operator on $\mathbb{R}^n$) such that*

- *each row and column of A contains at most L non-zero elements*
- *for each entry $A_{i,j}$, $|A_{i,j}| \leq M$*

then $\|A\| \leq 2LM$.

Proof. For unit-vectors $f, g \in \mathbb{R}^n$

$$\left|\langle A(f), g\rangle\right| = \left|\sum_{1\leq i,j\leq n} A_{i,j} f(i) g(j)\right| \leq M \sum_{1\leq i,j\leq n} \left|f(i)g(j)\right|.$$

That is

$$\left|\langle A(f), g\rangle\right| \leq \sum_{1\leq i,j\leq n} \left(f(i)^2 + g(j)^2\right).$$

Note that the number of occurrences of each $f(i)^2$ or $g(j)^2$ is at most L hence $\left|\langle A(f), g\rangle\right| \leq 2LM$. ∎

Now let us turn to part (b). The convergence of $\{\mu_n^i\}_{n=1}^{\infty}$ means that

$$\lim_{n\to\infty} \int_0^K P(t)\, d\mu_n^i(t)$$

exists for any real polynomial P. That is one needs to prove that

$$\lim_{n\to\infty} \frac{\sum_{j=1}^{|Q_i^n|} \left(\lambda_j^{i,n}\right)^r}{|Q_i^n|}$$

exists where $\{\lambda_j^{i,n}\}$ denotes the spectrum of the i-th Laplacian of Q^n. Hence it is enough to prove that the limit of normalized traces

$$\lim_{n\to\infty} \frac{\sum_{\sigma\in Q_i^n} \left(\Delta_{Q^n}^i(\sigma,\sigma)\right)^r}{|Q_i^n|} \tag{4}$$

exists. The value of $\Delta_{Q^n}^i(\sigma,\sigma)$ depends only on the r-neighboorhood of σ, therefore the convergence of the complexes $\{Q^n\}_{n=1}^{\infty}$ immediately implies the existence of the limit in (4).

Part (c) follows from the simple fact: If Q is a symmetric integer matrix then the product of its non-negative eigenvalues is an integer as well. Indeed, let $\lambda_1, \lambda_2, \ldots, \lambda_q$ the list of the non-zero eigenvalues of Q with multiplicities. Let $p(t) = \det(tI - Q)$ be the characteristic polynomial of Q. Then $p(t) = t^s q(t)$, where $q(0) \neq 0$. Obviously, q is an integer polynomial, and $|q(0)| = \left|\prod_{i=1}^{q} \lambda_q\right|$. ■

6. The Proof of Theorem 1

We need to prove the following lemma.

Lemma 6.1. *Let $\{K^n\}_{n=1}^{\infty} \subset \Sigma^d$ be simplicial complexes, then*

$$\lim_{n\to\infty} \frac{b^i(K_n)}{\left|V(K^n)\right|}$$

exists for any $i \geq 1$.

Indeed, by Lemma 6.1 we can immediately see that for any $\varepsilon > 0$ there exists $\delta > 0$ such that if $d_s(K, L) < \delta$ for some $K, L \in \Sigma^d$, then

$$\left| \frac{b^i(K)}{|V(K)|} - \frac{b^i(L)}{|V(L)|} \right| \leq \varepsilon, \tag{5}$$

for any $i \geq 1$. By the definition of the metric d_s there exists some $r \geq 1$ and $\rho > 0$ such that if

$$\left| p_M(\alpha) - p_N\alpha) \right| \leq \rho$$

for any $r' \leq r$ and $\alpha \in Z^{r',d}$ then $d_s(M, N) < \delta$ for any $M, N \in \Sigma^d$.

Now let $L^1, L^2, \ldots, L^m$ be a finite set of simplicial complexes such that for any finite simplicial complex K there exists $1 \leq j \leq m$ such that

$$\left| p_K(\alpha) - p_{L^j}(\alpha) \right| \leq \frac{\rho}{10} \tag{6}$$

for any $r' \leq r$ and $\alpha \in Z^{r',d}$, where r, ρ are the constants above. The existence of such finite system is clear from compactness.

By the classical Chernoff's inequality there exists $N_\varepsilon > 0$ such that the following holds:

Let $M \in \Sigma^d$ be an arbitrary simplicial complex. Pick N_ε random vertices of M and for any $r' \leq r$ and $\alpha \in Z^{r',d}$ let $Q(M, \alpha)$ be the number of picked vertices x such that $B_{r'}(x) \cong \alpha$. Then

$$\text{Prob}\left\{ \left| \frac{Q(M,\alpha)}{N_\varepsilon} - p_M(\alpha) \right| > \frac{\rho}{10} \text{ for at least one } \alpha \right\} < \varepsilon \tag{7}$$

Thus we have the following testing algorithm. Take the simplicial complex M as an input. Pick N_ε random vertices and calculate $Q(M, \alpha)$ for all $r' \leq r$, $\alpha \in Z^{r,d}$, where r is the constant above. Check the list $L^1, L^2, \ldots, L^m$. By (7), with probability more than $(1-\varepsilon)$ we find an L^j such that $\left| Q(M, \alpha) - P_{L_j}(\alpha) \right| < \frac{\rho}{5}$ for any α. Let $b^i(L_j)$ be our guess. Then by (5) with probability more than $1 - \varepsilon$

$$\left| \frac{b^i(L^j)}{|V(L^j)|} - \frac{b^i(M)}{|V(M)|} \right| < \varepsilon.$$

So, let us prove Lemma 6.1. By Proposition 5.1

$$\lim_{n\to\infty} \frac{\dim \operatorname{Ker} \Delta^i_{K^n}}{|K^n_i|}$$

exists. Also, by the definition of the convergence of simplicial complexes $\lim_{n\to\infty} \frac{|K^n_i|}{|V(K^n_i)|}$ exists, hence the lemma, and thus our Theorem follows. ∎

REFERENCES

[1] I. Benjamini and O. Schramm, Recurrence of distributional limits of finite planar graphs, *Electron. J. Probab.*, **6** (2001), no. 23, 13 pp. (electronic).

[2] G. Elek, Combinatorial Heat Kernels and Index Theorems, *Journal of Functional Analysis*, **129** (1995) no. 1, 64–79.

[3] G. Elek, *A Regularity Lemma for Bounded Degree Graphs and Its Applications: Parameter Testing and Infinite Volume Limits* (preprint), URL: http://arxiv.org/pdf/0711.2800v2.

[4] O. Goldreich and D. Ron, Property Testing in Bounded Degree Graphs, *Algorithmica*, **32** (2002), 302–343.

[5] W. Lück, Approximating L^2-invariants by their finite-dimensional analogues, *Geom. Funct. Anal.*, **4** (1994), no. 4, 455–481.

[6] W. Lück, L^2-torsion and 3-manifolds, *Low-dimensional topology (Knoxville, TN, 1992)*, 75–107. Conf. Proc. Lecture Notes Geom. Topology, III, Internat. Press (Cambridge, MA, 1994).

Gábor Elek

Alfréd Rényi Institute
H-1053 Reáltanoda u. 13–15.
Budapest
Hungary

e-mail: elek@renyi.hu

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 151–172.

Rigid and Globally Rigid Graphs with Pinned Vertices

TIBOR JORDÁN*

We consider rigid and globally rigid bar-and-joint frameworks (resp. graphs) in which some joints (resp. vertices) are pinned down and hence their positions are fixed. We give an overview of some old and new results of this branch of combinatorial rigidity with an emphasis on the related optimization problems.

In one of these problems the goal is to find a set P of vertices of minimum total cost for which the positions of all vertices become uniquely determined when P is pinned down. For this problem, which is motivated by the localization problem in wireless sensor networks, we give a constant factor approximation algorithm.

1. Introduction

A *bar-and-joint framework* (or simply *framework*) (G,p) in d-space is a graph $G=(V,E)$ and a map $p : V \to \mathbb{R}^d$. We also say that (G,p) is a d-dimensional *realization* of G. We can think of the edges and vertices of G in the framework as rigid (fixed length) bars and universal joints, respectively. An *infinitesimal motion* is a map $x : V \to \mathbb{R}^d$ satisfying

$$\big(p(v_i)-p(v_j)\big)\big(x(v_i)-x(v_j)\big)=0$$

for all edges $v_iv_j \in E$. The initial velocities obtained by differentiating a smooth motion of the (vertices of the) framework which preserves the edge lengths give rise to an infinitesimal motion of (G,p). The *rigidity matrix*

*This work was supported by the Mobile Innovation Centre, funded by the Hungarian National Office for Research and Technology.

of the framework (G,p) is the matrix $R(G,p)$ of size $|E| \times d|V|$, where, for each edge $e = v_iv_j \in E$, in the row corresponding to e, the entries in the two columns corresponding to vertices i and j contain the d coordinates of $\big(p(v_i) - p(v_j)\big)$ and $\big(p(v_j) - p(v_i)\big)$, respectively, and the remaining entries are zeros.

Example. The rigidity matrix of the framework of Figure 1 is as follows. The rows correspond to edges ab, bc, ca, cd, in this order, and consecutive pairs of columns correspond to vertices a, b, c, d.

$$\begin{pmatrix} 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 1 & -1 \end{pmatrix}$$

Thus x (viewed as a vector in $\mathbb{R}^{d|V|}$) is an infinitesimal motion if and only if $R(G,p)x = 0$. Each translation and rotation of $\mathbb{R}^d$ gives rise to a smooth motion of (G,p) and hence to an infinitesimal motion of (G,p). These rigid motions of $\mathbb{R}^d$ give rise to a subspace of dimension $\binom{d+1}{2}$ in the null-space of $R(G,p)$. Hence

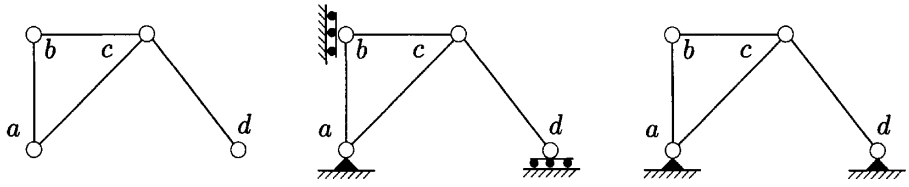
Lemma 1.1 [31, Lemma 11.1.3]. *Let (G,p) be a framework in $\mathbb{R}^d$. Then*

$$\operatorname{rank} R(G,p) \leq S(n,d), \tag{1}$$

where $n = |V(G)|$ and

$$S(n,d) = \begin{cases} nd - \binom{d+1}{2} & \text{if } n \geq d+2 \\ \binom{n}{2} & \text{if } n \leq d+1. \end{cases}$$

We say that a framework (G,p) is *infinitesimally rigid* in $\mathbb{R}^d$ if the rank of its rigidity matrix $R(G,p)$ is maximum, i.e. if equality holds in (1). A framework is *rigid* if it has no non-trivial smooth motions. Thus infinitesimal rigidity is a sufficient condition for rigidity. It is known that for "generic" frameworks the two notions are the same. We refer the reader to [10, 31, 32] for more details on the theory of rigid frameworks.



(a) (b) (c)

Fig. 1. A framework in $\mathbb{R}^2$ on four vertices (left). The coordinates of the vertices are as follows: $p(a) = (0,0)$, $p(b) = (0,1)$, $p(c) = (1,1)$, $p(d) = (2,0)$. Since $2|V| - \operatorname{rank} R(G,p) = 4$, to fix the framework one needs tracks of co-dimension four in total, which can be achieved by two one-dimensional tracks and a pin (middle) or two pins (right).

2. Rigid Frameworks with Pinned Vertices

Let $G = (V, E)$ be a graph and consider a d-dimensional realization (G, p) of G. We may fix (G, p) in $\mathbb{R}^d$ by restricting the infinitesimal motions of its vertices to given subspaces of $\mathbb{R}^d$. Suppose that for all vertices $v \in V$ we are given a subspace $U(v) \subseteq \mathbb{R}^d$, generated by a subset of the standard basis of $\mathbb{R}^d$. We call $U(v)$ the *track* of v and we say that (G, p) is *fixed* by the given set of tracks if the only infinitesimal motion x of (G, p) satisfying $x(v) \in U(v)$ for all $v \in V$ is the zero vector $x = 0$. In most cases we shall be interested in the special case when each track is either zero- or d-dimensional. We say that $P \subseteq V$ is a *pinning set* if (G, p) is fixed by the tracks $U(v) = \{0\}$ if $v \in P$, $U(v) = \mathbb{R}^d$ if $v \notin P$. We also say that the vertices in P are *pinned down,* or that each vertex of P is a *pin.*

The following lemma establishes the connection between tracks (pins) that fix a framework and its rigidity matrix (see also [26, Statement 8.2.1]). Note that each track $U(v)$ of dimension k, $0 \leq k \leq d$, corresponds naturally to a subset of size k of the d columns of the rigidity matrix which belong to v.

Lemma 2.1. *Let (G, p) be a framework in $\mathbb{R}^d$, let $U = (U(v) : v \in V)$ be a family of tracks, and let R_U be the matrix consisting of all columns of $R(G, p)$ which correspond to the tracks $U(v)$, $v \in V$. Then*

(i) *U fixes (G, p) if and only if the columns of R_U are linearly independent,*

(ii) *P is a pinning set if and only if the $d|V - P|$ columns of $R(G, p)$ indexed by $V - P$ are linearly independent.*

One may ask for an *optimal family of tracks* that fixes a given framework by using the least possible total restriction, i.e. an assignment $U = \big(U(v), v \in V\big)$ for which U fixes (G,p) and

$$\sum_{v \in V} \big(d - \dim U(v)\big)$$

is minimum. By Lemma 2.1(i) an optimal family of tracks is easy to find by using a greedy algorithm to identify a maximum size independent set of columns in $R(G,p)$. Furthermore, the optimum is unchanged if we restrict the matrix to a maximum size set of independent rows (or if we consider the corresponding subgraph of G). It is also clear that

$$\min\left\{\sum_{v \in V} \big(d - \dim U(v)\big) \ : \ U \text{ fixes } (G,p)\right\} = d|V| - \operatorname{rank} R(G,p).$$

We obtain a much more difficult problem if we impose restrictions on the dimension of the tracks. This is the case, for example, when we consider pinning sets. The *pinning number*, $\operatorname{pin}_d(G,p)$, of (G,p) is defined to be the size of a smallest pinning set for (G,p). For $d = 2$ Lemma 2.1(ii) implies that the smallest pinning set problem can be formulated as a matroid matching problem in a linearly represented matroid and hence $\operatorname{pin}_2(G,p)$ can be computed in polynomial time by using the algorithm of Lovász [21]. A combinatorial formula for $\operatorname{pin}_2(G,p)$ was also given by Lovász [22]. Mansfield [25] proved that the problem of computing $\operatorname{pin}_3(G,p)$ for a framework (G,p) is NP-hard.

On the other hand, a recent result of Szabó [30] shows that there exist tractable cases even in dimensions larger than two.

Theorem 2.2 [30]. *The following problem is polynomial time solvable. Given a framework (G,p) in $\mathbb{R}^3$ and a partition $V = V_1 \cup V_2$, find a family $U = \big(U(v) : v \in V\big)$ of tracks minimizing $\sum_{v \in V} \big(d - \dim U(v)\big)$ such that U fixes (G,p) and*

(i) $\dim U(v) \in \{0,1,3\}$ *for* $v \in V_1$, *and*

(ii) $\dim U(v) \in \{0,2,3\}$ *for* $v \in V_2$.

The algorithm in [30] also uses Lovász' matroid matching algorithm as a subroutine.

If we are also given a cost function on the vertices, we may look for a pinning set of minimum total cost. Baudis et al [2] proposed an approximation

algorithm for this more general problem. For d-dimensional frameworks the approximation guarantee is $1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{d}$. Their algorithm is based on a general result about minimum cost spanning sets in d-polymatroids, see also Section 8.

3. Rigid Graphs with Pinned Vertices

The rigidity matrix of a d-dimensional framework (G,p) defines the *rigidity matroid* of (G,p) on the ground set E where a set of edges $F \subseteq E$ is independent if and only if the rows of the rigidity matrix indexed by F are linearly independent. A framework (G,p) is *generic* if the set of coordinates of the points $p(v)$, $v \in V$, is algebraically independent over the rationals. Thus, since the entries of the rigidity matrix are polynomial functions with integer coefficients, any two generic d-dimensional frameworks (G,p) and (G,q) have the same rigidity matroid. We call this the *d-dimensional rigidity matroid* $\mathcal{R}_d(G)$ of the graph G. We denote the rank of $\mathcal{R}_d(G)$ by $r_d(G)$. We say that a graph $G=(V,E)$ is *generically infinitesimally rigid*, or simply *rigid*, in $\mathbb{R}^d$ if $r_d(G)=S(n,d)$. We say that a graph $G=(V,E)$ is *independent* in $\mathbb{R}^d$ if E is independent in $\mathcal{R}_d(G)$. It is not difficult to see that $\mathcal{R}_1(G)$ is the cycle matroid of G. It remains an open problem to find good characterizations for independence or, more generally, the rank function in the d-dimensional rigidity matroid of a graph when $d \geq 3$.

Similarly, any two generic d-dimensional frameworks on G have the same pinning number. Thus we may define the *pinning number of* G, $\mathrm{pin}_d(G)$, as the pinning number of (G,p) of any generic framework (G,p) in $\mathbb{R}^d$. It is easy to see that $\mathrm{pin}_d(G) \leq \mathrm{pin}_d(G,p)$ for all frameworks (G,p). The next lemma implies that computing the pinning number of G is the same as finding a smallest complete graph whose addition to G makes it rigid. For a set $P \subseteq V(G)$ let $G+K(P)$ denote the graph obtained from G by joining all pairs of non-adjacent vertices of P.

Lemma 3.1. *Let $G=(V,E)$ be a graph and $P \subseteq V$ with $|P| \geq d$. Let (G,p) be a generic realization of G in $\mathbb{R}^d$. Then P is a pinning set for (G,p) if and only if $G+K(P)$ is rigid in $\mathbb{R}^d$.*

Proof. Let $G'=G+K(P)$. First suppose that G' is rigid and consider the rigidity matrix $R(G',p)$. Since G' is rigid, the only solutions u to the equation $R(G',p)u=0$ are from rigid congruences of $\mathbb{R}^d$. Thus, since (G',p)

is generic, each non-zero solution leaves at most $(d-1)$ vertices fixed i.e. has at most $(d-1)$ zero entries. Suppose $R\big(G[V-P],p\big)$ has linearly dependent columns. Then we can find a non-zero solution u' to $R\big(G[V-P],p\big)\,u'=0$. By extending u' to u by putting 0 in the components corresponding to P we obtain a non-zero solution to $R(G',p)u=0$ with at least $|P|\geq d$ zeros, a contradiction. Thus P is a pinning set by Lemma 2.1(ii).

Now suppose that P is a pinning set and order the columns of $R=R(G',p)$ so that the columns of P come first and the rows of $E''=E\big(G'[P]\big)$ come first. (Then the upper right quarter is 0.) Hence $r(R)\geq r\big(R[P,E'']\big)+r\big(R[V-P,E-E'']\big)=d|P|-\binom{d+1}{2}+d|V-P|=d|V|-\binom{d+1}{2}$ (by using Lemma 2.1(ii) and that $G'[P]$ is rigid and $|P|\geq d$). Thus G' is rigid. ■

Next we show that in the pinning problem we may assume that G is independent.

Lemma 3.2. *Let $F\subseteq E$ be a maximal edge set of $G=(V,E)$ for which $H=(V,F)$ is independent in $\mathcal{R}_d$. Then*

(i) *each pinning set of G is a pinning set of H,*

(ii) $\mathrm{pin}_d(H)=\mathrm{pin}_d(G)$.

Proof. To prove (i) suppose, for a contradiction, that there exists a pinning set P of G for which $H+K(P)$ is not rigid. Since $G+K(P)$ is rigid, we have $r_d\big(G+K(P)\big)>r_d\big(H+K(P)\big)$, which implies that there is an edge $e\in E+E\big(K(P)\big)-\big(F+E\big(K(P)\big)\big)=E-F$ for which $F+e$ is independent, contradicting the maximality of F. This proves (i), from which (ii) follows immediately. ■

It follows from the observations above that the pinning problem in graphs (or in generic frameworks) can be attacked by purely combinatorial methods provided good characterizations for independent and rigid graphs are available. This is the case when $d=2$ and we shall discuss this approach in the 2-dimensional case in the forthcoming sections.

Mansfield [25] proved that the problem of computing $\mathrm{pin}_3(G)$ for a graph G is NP-hard (see also [7] for a different proof), so the pinning problem in higher dimensions seems untractable. The following related result, however, might be useful in a different context as it points to a connection between high connectivity and rigidity in 3-space. It may be considered as a first step towards the Lovász–Yemini conjecture [23], which asserts that sufficiently highly connected graphs are rigid in 3-space (and hence their pinning number is three).

Theorem 3.3 [14]. *Let $G = (V, E)$ be a 10-connected graph. Then* $\text{pin}_3(G) \leq \frac{3|V|}{4} + 4$.

4. The Two-dimensional Rigidity Matroid

In the rest of the paper we will be concerned with the case when $d = 2$ and suppress the subscript d accordingly. In this section we first describe the characterization of independent and rigid graphs and prove some additional structural results which may also be useful in the solution of the pinning problem. For $X \subseteq V$ let $E_G(X)$ denote the set, and $i_G(X)$ denote the number of edges in $G[X]$, that is, in the subgraph induced by X in G. We say that a graph G is *sparse* if $i_G(X) \leq 2|X| - 3$ for all $X \subseteq V$ with $|X| \geq 2$. It is easy to show, by using the 2-dimensional case of Lemma 1.1, that independent graphs are sparse. Laman [18] proved that this necessary condition is also sufficient.

Theorem 4.1 [18]. *A graph $G = (V, E)$ is independent if and only if G is sparse.*

A *cover* of $G = (V, E)$ is a collection $\mathcal{X} = \{X_1, X_2, \ldots, X_t\}$ of subsets of V, each of size at least two, such that $\cup_1^t E(X) = E$. The cover is said to be *thin* if $|X_i \cap X_j| \leq 1$ for all $i \neq j$. The *value* $\text{val}(\mathcal{X})$ of the cover is $\sum_{i=1}^t (2|X_i| - 3)$.

Let $\mathcal{X}$ be a thin cover of G and let $F \subseteq E$ be a set of edges for which $H = (V, F)$ is sparse. Then we have $|F \cap E_G(X_i)| \leq 2|X_i| - 3$ for all $1 \leq i \leq t$. Thus

$$|F| \leq \text{val}(\mathcal{X}). \tag{2}$$

We define a *rigid component* of a graph $G = (V, E)$ to be a maximal rigid subgraph of G. By the "plane gluing lemma" (see [31, Lemma 3.1.4]), which says that the union of two rigid graphs with at least two vertices in common is rigid, it follows that the vertex sets of the rigid components form a thin cover of G. In a sparse graph H we call a set $X \subseteq V(H)$ *critical* if $i_H(X) = 2|X| - 3$ holds. It follows from the gluing lemma that if $X, Y \subset V(H)$ are critical sets in H with $|X \cap Y| \geq 2$ then $X \cup Y$ is also critical (see also [13, Lemma 2.3]).

Lemma 4.2. *Let $G = (V, E)$ be a graph, let $F \subseteq E$ be a maximal edge set in G for which $H = (V, F)$ is sparse. Then the family $\mathcal{X} = \{X_1, X_2, \ldots, X_t\}$ of maximal critical sets in H satisfies that*

(a) *$\mathcal{X}$ is a thin cover of G with $|F| = \mathrm{val}(\mathcal{X})$,*

(b) *$\mathcal{X}$ is equal to the family of vertex sets of the rigid components of G.*

Proof. (a) The maximality of the critical sets implies that $|X_i \cap X_j| \leq 1$ for all $1 \leq i < j \leq t$. Since every single edge of F induces a critical set, it follows that $\mathcal{X} = \{X_1, X_2, \ldots, X_t\}$ is a thin cover of H. Thus

$$|F| = \sum_1^t |E_H(X_i)| = \sum_1^t (2|X_i| - 3).$$

To complete the proof we show that $\mathcal{X}$ is a cover of G as well. Choose $uv \in E - F$. Since F is a maximal sparse subset of E, $F + uv$ is not sparse. Thus there exists a set $X \subseteq V$ such that $u, v \in X$ and $i_H(X) = 2|X| - 3$. Hence X is a critical set in H. This implies that $X \subseteq X_i$ and hence $uv \in E_G(X_i)$ for some $1 \leq i \leq t$.

(b) Clearly, $G[X_i]$ is rigid for all $1 \leq i \leq t$ by Theorem 4.1. Suppose that $H[C]$ is not critical, where C is the set of vertices of some rigid component of G. Thus $|J| \leq 2|C| - 4$, where $J = E\big(H[C]\big)$. Since $G[X_i]$ is rigid, it follows from the gluing lemma that $\mathcal{X}' = \big\{X_i \in \mathcal{X} : |X_i \cap C| \geq 2\big\}$ is a thin cover of $G[C]$ with $|J| = \mathrm{val}(\mathcal{X}')$. Thus we can use (2) to deduce that for any subset $F' \subseteq E\big(G[C]\big)$ which induces a sparse subgraph on vertex set C we have $|F'| \leq \mathrm{val}(\mathcal{X}') = |J| \leq 2|C| - 4$, contradicting the fact that $G[C]$ is rigid. This completes the proof of (b). ∎

Lemma 4.2(a) shows that for the edge set of a maximal sparse subgraph and for its maximal critical sets we have equality in (2). This implies the following rank formula of the rigidity matroid, due to Lovász and Yemini, and shows that a maximum size sparse edge set can be found greedily.

Theorem 4.3 [23]. *Let $G = (V, E)$ be a graph. Then*

$$r(G) = \min\big\{\mathrm{val}(\mathcal{X}) : \mathcal{X} \text{ is a thin cover of } G\big\}.$$

A maximum size sparse edge set (and the rigid components of a graph) can be found in time $O(n^2)$, see e.g. [4].

We may simplify the min-max formula of Theorem 4.3 when the graph is obtained from a sparse graph by 'pinning' a set of vertices. For a set $X \subseteq V$ let $e(X)$ denote the number of edges with at least one end-vertex in X.

Lemma 4.4. *Suppose that $G = (V, E)$ is a sparse graph and let $P \subseteq V$ with $|P| \geq 2$. Let $G' = G + K(P)$. Then*

$$r(G') = \min_{P \subseteq Z} 2|Z| - 3 + e(V - Z).$$

Proof. Let $Z \subseteq V$ with $P \subseteq Z$ and consider the thin cover $\mathcal{Z} = \big\{Z \cup \big\{\{u, v\} : uv \in E - E(Z)\big\}\big\}$ of G'. Then $r(G') \leq \mathrm{val}(\mathcal{Z}) = 2|Z| - 3 + e(V - Z)$.

To see that equality holds for some $Z \subseteq V$ choose a maximal edge set F in G' for which $H = (V, F)$ is sparse and P is a critical set in H. Such an F can be constructed by extending the edge set F' of a minimally rigid subgraph of the complete graph $G'[P]$. Let $\mathcal{X}$ be the family of maximal critical sets of H. By Lemma 4.2(a) and Theorem 4.3 we have $r(G') = \mathrm{val}(\mathcal{X})$. Since P is critical in H, there is a set $Z \in \mathcal{X}$ with $P \subseteq Z$. Thus, since G is sparse and all edges of $K(P)$ are covered by Z, we have $i_G(X) = i_H(X) = 2|X| - 3$ for all $X \in \mathcal{X} - Z$. Hence $r(G') = \mathrm{val}(\mathcal{X}) = 2|Z| - 3 + \sum_{X \in \mathcal{X} - Z} i_G(X) = 2|Z| - 3 + e(V - Z)$, which completes the proof. ■

5. Optimal Families of Tracks and Smallest Pinning Sets

Let $G = (V, E)$ be a graph. First consider the problem of finding an optimal family of tracks, $U = \big(U(v) : v \in V\big)$, which fixes (G, p) for a generic realization of G in $\mathbb{R}^2$. As we have observed earlier, we may assume that G is independent (or equivalently, that G is sparse). Thus $|E| = 2|V| - k$ for some integer $k \geq 3$. It is also clear that $\sum_{v \in V} \big(2 - \dim U(v)\big) = k$ for an optimal family of tracks. The following algorithm, due to Lee et al. [19], determines an optimal family of tracks in $O(n^2)$ time. It uses $k - 2$ one-dimensional tracks (also called *sliders*) and one pin to fix (G, p). (For the remaining vertices the tracks are two-dimensional.)

The algorithm works as follows. First identify the rigid components of G. Mark one of the components, say C, as the *base*. For some edge uv in C assign a pin to u and a slider to v. This fixes the base. Then repeat the following until one rigid component remains: pick an edge ij which leaves the base and assign a slider to j. Update G by adding a new edge jk, where

ik is an edge in the base. Replace the base C by the rigid component of the updated graph containing uv.

The correctness of this algorithm follows from the fact that if C' is a rigid component that shares vertex i with C then the only motion of C' with respect to C is a rotation about i. Since the framework is generic, assigning a slider to j eliminates this motion and hence the distance between j and k becomes fixed. Thus every iteration increases the rank by one and therefore the algorithm will terminate with a rigid graph after adding at most $k-3$ sliders (not counting the pin and the slider added to fix the original base). The algorithm, when applied to (a generic realization of) the graph of Figure 1(a), may give the family of tracks shown by Figure 1(b).

We remark that combinatorial characterizations for the generic rigidity of bar-and-slider frameworks (which are bar-and-joint frameworks equipped with sliders at given joints) have been given in [19], and also in [17], where the authors consider the version in which the directions of the slider lines are also given.

Next consider the pinning problem.

Lemma 5.1. [7] *Let $G=(V,E)$ be a sparse graph and let $P \subseteq V$ with $|P| \geq 2$. Then P is a pinning set for G if and only if $2|X| \leq e(X)$ for all $X \subseteq V-P$.*

Proof. Suppose, for a contradiction, that P is a pinning set and $2|X| > e(X)$ for some $X \subseteq V-P$ and let $Z=V-X$. Then $\mathcal{X}=\{Z \cup \{\{u,v\} : uv \in E-E(Z)\}\}$ is a thin cover of $G+K(P)$ with $\mathrm{val}(\mathcal{X}) \leq 2|Z|-3+e(X) < 2|V|-3$. Thus, by Theorem 4.3, $G+K(P)$ is not rigid. Hence P is not a pinning set by Lemma 3.1, a contradiction.

Now suppose $2|X| \leq e(X)$ for all $X \subseteq V-P$. It follows from Lemma 4.4 that there is a thin cover $\mathcal{X}$ of $G+K(P)$ with $P \subseteq Z$ for some $Z \in \mathcal{X}$ and $r\big(G+K(P)\big) = \mathrm{val}(\mathcal{X}) = 2|Z|-3+e(V-Z)$. Since $e(V-Z) \geq 2|V-Z|$ this gives $\mathrm{val}(\mathcal{X}) = 2|V|-3$. Hence $G+K(P)$ is rigid and, by Lemma 3.1, P is a pinning set. ■

Thus finding a smallest pinning set is equivalent to finding a largest set $Y \subseteq V$ for which $e(X) \geq 2|X|$ for all $X \subseteq Y$. This can be formulated as a matching problem in an auxiliary graph (see Fekete [7]) and can be solved in $O(n^2)$ time. Fekete [7] also provides a min-max formula for $\mathrm{pin}_2(G)$. Makai and Szabó [24] deduce this formula by using polymatroidal methods.

We note that Servatius, Shai, and Whiteley [28] consider a different version of the pinning problem and provide a characterization and a decomposition result for the so-called pinned isostatic graphs.

6. The Network Localization Problem

In the network localization problem the locations of some nodes (called anchors) of a network as well as the distances between some pairs of nodes are known, and the goal is to determine the location of all nodes. This is one of the fundamental algorithmic problems in the theory of wireless sensor networks and has been the focus of a number of recent research articles and survey papers, see for example [1].

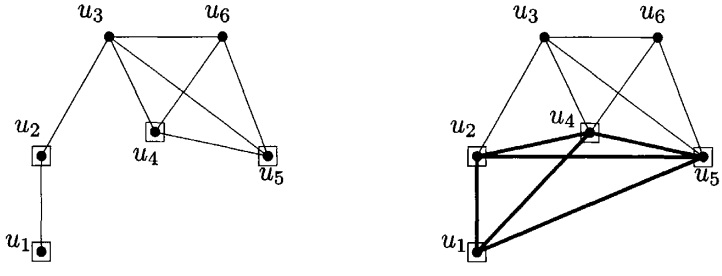


Fig. 2. The distance graph and the grounded graph of a network on six nodes, including four anchor nodes. The anchor nodes are in boxes. The network is uniquely localizable since it has at least three anchors and its grounded graph is globally rigid. This is a smallest anchor set which can guarantee unique localizability for the given set of distances.

A natural additional question is whether a solution to the localization problem is unique. The network, with the given locations and distances, is said to be *uniquely localizable* if there is a unique set of locations consistent with the given data. As we shall see, the unique localizability of a two-dimensional network, whose nodes are in generic position, can be characterized by using results from graph rigidity theory. In this case unique localizability depends only on the combinatorial properties of the network: it is determined completely by the *distance graph* of the network and the set of anchors, or equivalently, by the *grounded graph* of the network and the number of anchors. The vertices of the distance and grounded graph correspond to the nodes of the network. In both graphs two vertices are connected by an edge if the corresponding distance is explicitly known. In the grounded graph we have additional edges: all pairs of vertices corresponding

to anchor nodes are adjacent. See Figure 2. The grounded graph represents all known distances, since the distance between two anchors is determined by their locations. Before stating the basic observation about unique localizability we need some additional terminology. It is convenient to investigate localization problems with distance information by using frameworks, the central objects of rigidity theory.

Two frameworks (G,p) and (G,q) are *equivalent* if corresponding edges have the same lengths, that is, if $\|p(u)-p(v)\| = \|q(u)-q(v)\|$ holds for all pairs u, v with $uv \in E$, where $\|.\|$ denotes the Euclidean norm in $\mathbb{R}^d$. Frameworks (G,p), (G,q) are *congruent* if $\|p(u)-p(v)\| = \|q(u)-q(v)\|$ holds for all pairs u,v with $u,v \in V$. This is the same as saying that (G,q) can be obtained from (G,p) by an isometry of $\mathbb{R}^d$. We shall say that (G,p) is *globally rigid,* or that (G,p) is a *unique realization* of G in $\mathbb{R}^d$, if every framework which is equivalent to (G,p) is congruent to (G,p). We say that a graph G is *globally rigid* in $\mathbb{R}^d$ if every (or equivalently, if some) generic realization of G in $\mathbb{R}^d$ is globally rigid.

The next observation shows that unique localizability and global rigidity are, in some sense, the same.

Lemma 6.1 [1, 29]. *Let N be a network in $\mathbb{R}^d$ consisting of m anchors located at positions $p_1,\ldots,p_m$ and $n-m$ ordinary nodes located at $p_{m+1},\ldots,p_n$. Suppose that there are at least $d+1$ anchors in general position. Let G be the grounded graph of N and let $p=(p_1,\ldots,p_n)$. Then the network is uniquely localizable if and only if (G,p) is globally rigid.*

Globally rigid graphs in $\mathbb{R}^2$ have been characterized by Jackson and Jordán [13], relying on earlier results of Hendrickson [11] and Connelly [6]. We say that a graph G is *redundantly rigid* in $\mathbb{R}^2$ if $G-e$ is rigid in $\mathbb{R}^2$ for all $e \in E(G)$.

Theorem 6.2 [13]. *Let (G,p) be a generic framework in $\mathbb{R}^2$. Then (G,p) is globally rigid if and only if G is a complete graph on at most three vertices or G is 3-connected and redundantly rigid.*

Theorem 6.2 implies that global rigidity is indeed a generic property. It also implies that global rigidity can be tested in $O(n^2)$ time.

We shall consider the *minimum cost anchor set problem* in which the goal is, given the set of known distances in a network and a cost function on the nodes, to designate a minimum cost set of anchor nodes which makes the network uniquely localizable. Lemma 6.1 and Theorem 6.2 imply that

for generic networks we may reformulate the above problem in the following purely combinatorial form:

Given a graph $G = (V, E)$ and a function $c : V \to \mathbb{R}_+$, find a set $P \subseteq V$, $|P| \geq 3$, for which $G + K(P)$ is 3-connected and redundantly rigid, and $c(P) = \sum_{v \in P} c(v)$ is minimum.

In the next sections first we shall show that a relaxed version (in which the requirement is that the rigidity matroid of $G+K(P)$ must be connected) can be formulated as a matroid optimization problem. Then, based on this formulation, we shall develop a polynomial time approximation algorithm for the minimum cost anchor set problem. Note that the complexity status of each of the above problems is still open.

7. Graphs with a Connected Rigidity Matroid

Given a matroid $\mathcal{M} = (E, \mathcal{I})$, we define a relation on E by saying that $e, f \in E$ are related if $e = f$ or if there is a circuit C in $\mathcal{M}$ with $e, f \in C$. It is well-known that this is an equivalence relation. The equivalence classes are called the *components* of $\mathcal{M}$. If $\mathcal{M}$ has at least two elements and only one component then $\mathcal{M}$ is said to be *connected.*

We say that a graph $G = (V, E)$ is *M-connected* if $\mathcal{R}(G)$ is connected. For example, $K_{3,m}$ is M-connected for all $m \geq 4$. The *M-components* of G are the subgraphs of G induced by the components of $\mathcal{R}(G)$. It is easy to see that the M-components are pairwise edge-disjoint induced subgraphs. Theorem 6.2 and the following result show that M-connectivity is in between redundant rigidity and global rigidity.

Theorem 7.1 [13]. *Let G be a graph. Then*

(a) *if G is M-connected then G is redundantly rigid, and*

(b) *if G is 3-connected and redundantly rigid then G is M-connected.*

Since the M-components of G are redundantly rigid by Theorem 7.1, the partition of $E(G)$ given by the M-components is a refinement of the partition given by the rigid components, see Figure 3. The rigidity matroid of a graph G is the direct sum of the rigidity matroids of either the rigid components of G or the M-components of G. Furthermore, the vertex sets of the components in each of the above decompositions form a thin cover

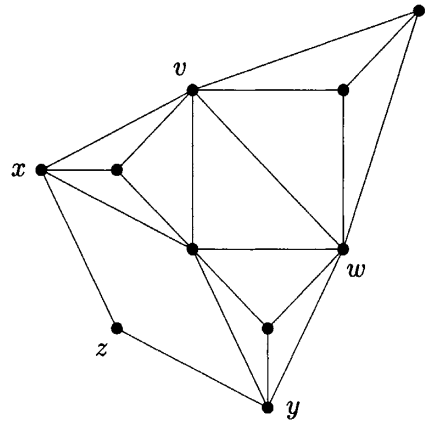


Fig. 3. This graph is rigid so has exactly one rigid component. It has five M-connected components: each of the three copies of K_4, and the remaining two copies of K_2.

of G with minimum value. This minimum value is equal to the rank of $\mathcal{R}(G)$ by Theorem 4.3.

The following lemma is easy to prove by standard matroid techniques.

Lemma 7.2. *Let $\mathcal{M} = (E, r)$ be a matroid on ground set E with rank function r and let $E_1, E_2, \ldots, E_t$ be the components of $\mathcal{M}$. Then*
(i) $r(E) = \sum_1^t r(E_i)$, *and*
(ii) *if $r(E) = \sum_1^q r(F_i)$ for some partition $F_1, F_2, \ldots, F_q$ of E and E_i is a component of $\mathcal{M}$ for some $1 \leq i \leq t$, then $E_i \subseteq F_j$ for some $1 \leq j \leq q$.*

The next lemma shows how this general result can be formulated in terms of subgraphs and covers in the special case when the matroid is the rigidity matroid of a graph. We say that a cover is *non-trivial* if it contains at least two sets.

Lemma 7.3 [9]. *$G = (V, E)$ is M-connected if and only if* $\mathrm{val}(\mathcal{X}) \geq 2|V| - 2$ *for all non-trivial covers $\mathcal{X}$ of G.*

Proof. First suppose that G is M-connected. Then G is rigid, and hence $\mathrm{val}(\mathcal{X}) \geq 2|V| - 3$ for all covers $\mathcal{X}$ of G by (the easy direction of) Theorem 4.3. Suppose that $\mathrm{val}(\mathcal{X}) = 2|V| - 3$ for some non-trivial cover $\mathcal{X} = \{X_1, X_2, \ldots, X_q\}$ of G. Let $F_i = E\big(G[X_i]\big)$, $1 \leq i \leq q$. We have $r(F_i) = 2|X_i| - 3$ for all $1 \leq i \leq q$, as $\mathcal{X}$ is a cover of G which minimizes $\mathrm{val}(\mathcal{X})$. Thus $r(E) = \mathrm{val}(\mathcal{X}) = \sum_1^q r(F_i)$, which contradicts Lemma 7.2(ii).

To prove the other direction suppose that $\mathrm{val}(\mathcal{X}) \geq 2|V| - 2$ for all non-trivial covers $\mathcal{X}$ of G, but G is not M-connected. Let $H_1, H_2, \ldots, H_t$ be the

M-components of G. Lemma 7.2(i) now implies that $2|V| - 3 \geq r(E) = \sum_1^t r\big(E(H_i)\big) = \sum_1^t \big(2|V(H_i)| - 3\big)$. Thus, since each edge of G belongs to some M-component and $t \geq 2$, $\mathcal{X} = \big\{V(H_1), V(H_2), \ldots, V(H_t)\big\}$ is a non-trivial cover of G with $\mathrm{val}\,(\mathcal{X}) \leq 2|V| - 3$. This contradicts our assumption. ■

7.1. M-connected graphs with pinned vertices

In the *M-connected pinning problem* the goal is to find a (smallest) set $P \subseteq V$ for which $G+K(P)$ is M-connected. The following lemma establishes the connection between the feasible solutions of the M-connected pinning problem and the M-components of G.

Lemma 7.4 [9]. *Let $G = (V, E)$ be a graph, let $\mathcal{H} = \{H_1, H_2, \ldots, H_t\}$ be the M-components of G, and let $P \subseteq V$ with $|P| \geq 4$. Then $G + K(P)$ is M-connected if and only if*

$$2|V| - 2 \leq 2|Z| - 3 + \sum_{H_i \in \mathcal{H}_Z} \big(2|V(H_i)| - 3\big) \tag{3}$$

holds for all $Z \subset V$ with $P \subseteq Z$, $Z \neq V$, where $\mathcal{H}_Z = \big\{H_i \in \mathcal{H} : V(H_i) \cap (V - Z) \neq \emptyset\big\}$.

Proof. First suppose that $G+K(P)$ is M-connected. Since every edge of G belongs to an M-component of G and $P \subseteq Z$, it follows that $\{Z\} \cup \big\{V(H_i) : H_i \in \mathcal{H},\ V(H_i) \cap (V - Z) \neq \emptyset\big\}$ is a cover of $G + K(P)$. This cover is non-trivial, since $Z \neq V$. Thus (3) follows from Lemma 7.3.

To prove the other direction suppose, for a contradiction, that (3) holds but $G' = G+K(P)$ is not M-connected. Let $\mathcal{H}' = \{H'_1, H'_2, \ldots, H'_q\}$ denote the M-components of G'. Since complete graphs on at least four vertices are M-connected, and $|P| \geq 4$, it follows that $G'[P]$ is M-connected. Thus there is an M-component of G', say H'_1, for which $P \subseteq V(H'_1)$. Let $Z' = V(H'_1)$ and $\mathcal{H}_{Z'} = \big\{H_i \in \mathcal{H} : V(H_i) \cap (V - Z') \neq \emptyset\big\}$. Note that $Z' \neq V$.

Claim 7.5. *Let $X \subseteq V$ be a set of vertices. Then $X = V(H'_j)$ for some M-component H'_j of G' with $2 \leq j \leq q$ if and only if $X = V(H)$ for some $H \in \mathcal{H}_{Z'}$.*

Proof. First consider an M-component $H'_j \in \mathcal{H}'$ with $j \geq 2$ and let $X = V(H'_j)$. Since $P \subseteq Z'$ and H'_1 is an induced subgraph of G' which has no edge in common with H'_j, it follows that $G[X]$ is M-connected and $X \cap (V - Z') \neq \emptyset$. Thus $X = V(H)$ for some $H \in \mathcal{H}_{Z'}$.

Next consider an M-component $H_i \in \mathcal{H}_{Z'}$ of G and put $X = V(H_i)$. $G'[X]$ is clearly M-connected. For a contradiction suppose that there is an M-component H'_j of G' with $V(H'_j) = Y \subseteq V$ for which X is a proper subset of Y. Then $|Y \cap Z'| \geq |Y \cap P| \geq 2$ must hold. Since $X \cap (V - Z') \neq \emptyset$, we have $j \geq 2$. This contradicts the fact that the M-components of G' are pairwise edge-disjoint. Thus $G'[X]$ is an M-component of G', which completes the proof. ■

By using Claim 7.5 and Lemma 7.2(i), and by applying (3) with $Z = Z'$, we obtain

$$2|V| - 3 \geq r(G') = 2\big|V(H'_1)\big| - 3 + \sum_{H_i \in \mathcal{H}_{Z'}} \big(2\big|V(H_i)\big| - 3\big) \geq 2|V| - 2,$$

a contradiction. ■

Let $G = (V, E)$ be a graph and let $\mathcal{H} = \{H_1, H_2, \ldots, H_t\}$ be the M-components of G. Let $H(G) = (V, \mathcal{E})$ be the hypergraph which contains $2\big|V(H_i)\big| - 3$ copies of the hyperedge $V(H_i)$ for each $H_i \in \mathcal{H}$, $1 \leq i \leq t$. Note that since the M-components are rigid it follows from Lemma 7.2(i) that $|\mathcal{E}| = r(G) \leq 2|V| - 3$. By letting $Y = V - Z$ in Lemma 7.4 and using the above definitions we obtain:

Lemma 7.6. *Let $G = (V, E)$ be a graph, let $\mathcal{H} = \{H_1, H_2, \ldots, H_t\}$ be the M-components of G, and let $P \subseteq V$ with $|P| \geq 4$. Then $G + K(P)$ is M-connected if and only if*

$$2|Y| + 1 \leq e_{H(G)}(Y) \tag{4}$$

holds for all non-empty subsets $Y \subseteq V - P$, where $e_{H(G)}(Y)$ denotes the number of hyperedges $e \in \mathcal{E}$ with $e \cap Y \neq \emptyset$.

A hypergraph $F = (V, \mathcal{F})$ satisfying $|\cup \mathcal{F}'| \geq |\mathcal{F}'| + 1$ for all $\emptyset \neq \mathcal{F}' \subseteq \mathcal{F}$ is called a *hyperforest*. Inequality (4) can be reformulated in terms of hyperforests as follows. Let $L(G) = (W, \mathcal{U})$ be the hypergraph obtained from the dual hypergraph of $H(G)$ by duplicating every hyperedge[1]. For a

[1] Consider the bipartite incidence graph G^* of $H(G)$ and split each vertex $u \in V$ into two vertices u', u''. See Figure 4. Then we obtain the bipartite incidence graph of $L(G)$ by interchanging the color classes.

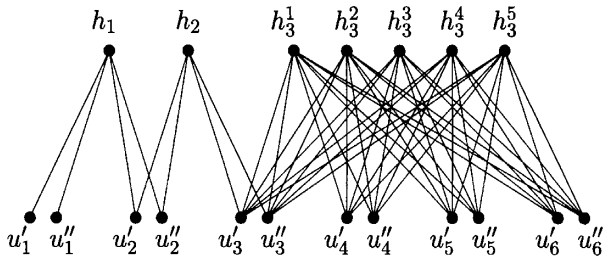


Fig. 4. The bipartite incidence graph of $L(G)$, where G is the graph of Figure 2.

set $X \subseteq V$ let $\mathcal{U}(X)$ denote the set of hyperedges corresponding to X in $L(G)$. Thus $\big|\mathcal{U}(X)\big| = 2|X|$.

Lemma 7.7. *Let $G = (V, E)$ be a graph and let $P \subseteq V$ with $|P| \geq 4$. Then P satisfies (4) if and only if $\mathcal{U}(V - P)$ is a hyperforest.*

Lorea [20] proved that the edge sets of the subhypergraphs of a hypergraph H' which are hyperforests form the family of independent sets of a matroid. A matroid arising this way is called the *circuit matroid* of the hypergraph H' and will be denoted by $\mathcal{M}_{H'}$. We call a matroid which is the circuit matroid of a hypergraph a *hypergraphic matroid.* Let $\mathcal{M}$ be a matroid on ground-set S and suppose that S is partitioned into a set A of pairs. A subset $M \subseteq A$ is a *matroid matching* if the union of the pairs in M is independent in $\mathcal{M}$. In the *matroid matching problem* the goal is to find a largest matroid matching, see [27, Chapter 43]. Lovász [21] has shown that this problem may require exponential time in general but can be solved polynomially if the matroid is represented by a set of vectors in some linear space.

By the above discussion and Lemma 7.7 it follows that the problem of finding a smallest set P for which $G + K(P)$ is M-connected can be formulated as finding a largest matroid matching in the hypergraphic matroid $\mathcal{M}_{L(G)}$, in which the doubled hyperedges form the pairs. Hypergraphic matroids are known to be linear, but it is not known how to find a suitable linear representation. The complexity status of the matroid matching problem in hypergraphic matroids is still open. Nevertheless, this formulation can be used to design a randomized algorithm, see [8], or a constant factor approximation algorithm which works for the more difficult minimum cost version as well.

To describe the approximation algorithm we need the following concepts. A *2-polymatroid* is a pair (S, f), where S is a finite ground set and f is a non-negative, monotone increasing, integer-valued, and submodular function on

the subsets of S, for which $f(s) \leq 2$ for all $s \in S$. A set $X \subseteq S$ is *spanning* if $f(X) = f(S)$.

Let $G = (V, E)$ be a graph and $X \subseteq V$. Let us define $b : 2^V \to \mathbb{Z}_+$ by letting

$$b(X) = r^*\big(\mathcal{U}(X)\big), \tag{5}$$

where r^* is the rank function of the matroid dual of the hypergraphic matroid $\mathcal{M}_{L(G)}$. Then (V, b) is a 2-polymatroid.

For a spanning set $X \subseteq V$ we have $r^*\big(\mathcal{U}(X)\big) = b(X) = b(V) = r^*\big(\mathcal{U}(V)\big)$. Thus X is spanning if and only if the set corresponding to $\mathcal{U}(V - X)$ is independent in $\mathcal{M}_{L(G)}$. Together with Lemmas 7.6 and 7.7 this implies:

Lemma 7.8. *Let $G = (V, E)$ be a graph and $P \subseteq V$ with $|P| \geq 4$. Then $G + K(P)$ is M-connected if and only if P is a spanning set of the 2-polymatroid (V, b).*

8. Low Cost Anchor Sets in Uniquely Localizable Networks

Given a 2-polymatroid (S, f) and a cost function $c : S \to \mathbb{R}$, the *minimum cost spanning set problem* is to find a spanning set X of the 2-polymatroid that minimizes $c(X) = \sum_{s \in X} c(s)$. Baudis et al. [2] verified that the GSS (Greedy Spanning Set) algorithm is a constant factor approximation algorithm for this problem. Algorithm GSS starts with $X = \emptyset$ and, as long as $f(X) < f(S)$ holds, adds a new element s to X for which $\frac{c(s)}{f(X+s)-f(X)}$ is minimum.

Theorem 8.1 [2]. *Let (S, f) be a 2-polymatroid, let $c : S \to \mathbb{R}$ be a cost function, and let X_{opt} be a spanning set of minimum cost. Then*

$$c(X) \leq \frac{3}{2} c(X_{\text{opt}}),$$

where X is the spanning set output by algorithm GSS.

Lemma 7.8 and Theorem 8.1 give rise to a $\frac{3}{2}$-approximation algorithm for the *minimum cost M-connected pinning problem.* To see this it remains

to note that by using bipartite matching algorithms it is easy to test independence in $\mathcal{M}_{L(G)}$ and evaluate $b(X)$ for some $X \subseteq V$ in polynomial time.

To obtain an approximation algorithm for the minimum cost anchor set problem (defined in Section 6) we also need a subroutine for the *minimum cost 3-connected pinning problem.* Let $H = (V, E)$ be a 2-connected graph. For some $X \subseteq V$ let $N(X)$ denote the set of neighbours of X. We say that $X \subset V$ is *tight* if $|N(X)| = 2$ and $X \cup N(X) \neq V$. The following lemma shows that a minimum cost set P' for which $H + K(P')$ is 3-connected can be found, in a greedy manner, in linear time.

Lemma 8.2. *Let $H = (V, E)$ be 2-connected and let $P' \subseteq V$. Then $H + K(P')$ is 3-connected if and only if $P' \cap X \neq \emptyset$ for all minimal tight sets X of H. Furthermore, the minimal tight sets of H are pairwise disjoint and can be found in linear time.*

Recall that redundant rigidity and M-connectivity are the same for 3-connected graphs by Theorem 7.1. Thus, by combining the approximation algorithm for the minimum cost M-connected pinning problem and the algorithm for the minimum cost 3-connected pinning problem we obtain a constant factor approximation algorithm for the minimum cost anchor set problem.

Theorem 8.3. *There is a polynomial time $\frac{5}{2}$-approximation algorithm for the minimum cost anchor set problem.*

Proof (sketch). Let c^* denote the optimum value. By checking all feasible solutions $P \subseteq V$ with $|P| = 3$ we may suppose that the optimal solution has at least four vertices. First we compute a close-to-optimal solution P for the minimum cost M-connected pinning problem with $c(P) \leq \frac{3}{2}c^*$. Since $G' = G + K(P)$ is M-connected, it is 2-connected. Then we compute an optimal solution P' for the minimum cost 3-connected pinning problem on G'. Clearly, $c(P') \leq c^*$. It is also clear that $G + K(P \cup P')$ is 3-connected and M-connected. Furthermore, $c(P \cup P') \leq c(P) + c(P') \leq \frac{5}{2}c^*$ holds. ■

We remark that the above methods can be used to design a constant factor approximation algorithm for the corresponding *augmentation problem* as well, in which the goal is to add a smallest set F of new edges to G such that $G + F$ is globally rigid. We omit the details.

Acknowledgement. I thank Zsolt Fekete and Bill Jackson for several useful comments.

REFERENCES

[1] J. Aspnes, T. Eren, D. K. Goldenberg, A. S. Morse, W. Whiteley, Y. R. Yang, B. D. O. Anderson and P. N. Belhumeur, A theory of network localization, *IEEE Trans. on Mobile Computing,* vol. 5, issue 12 (2006), pp. 1663–1678.

[2] G. Baudis, C. Gröpl, S. Hougardy, T. Nierhoff and H. J. Prömel, *Approximating minimum spanning sets in hypergraphs and polymatroids,* Tech. Report, Humboldt-Universität zu Berlin, 2000.

[3] A. R. Berg and T. Jordán, A proof of Connelly's conjecture on 3-connected circuits of the rigidity matroid, *J. Combinatorial Theory, Ser. B.,* Vol. **88** (2003), 77–97.

[4] A. R. Berg and T. Jordán, Algorithms for graph rigidity and scene analysis, *Proc. 11th Annual European Symposium on Algorithms (ESA),* 2003, (G. Di Battista, U. Zwick, eds.), Springer Lecture Notes in Computer Science 2832 (2003), pp. 78–89.

[5] R. Connelly, On generic global rigidity, Applied geometry and discrete mathematics, 147–155, *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.,* 4, Amer. Math. Soc., Providence, RI (1991).

[6] R. Connelly, Generic global rigidity, *Discrete Comput. Geom.,* **33** (2005), no. 4, 549–563.

[7] Z. Fekete, Source location with rigidity and tree packing requirements, *Operations Research Letters,* **34**, Issue 6 (2006), pp. 607–612.

[8] Z. Fekete and T. Jordán, Uniquely localizable networks with few anchors, *Proc. 4th Japanese-Hungarian Symposium on Discrete Mathematics and Its Applications,* Budapest (June 2005), pp. 144–148.

[9] Z. Fekete and T. Jordán, Uniquely localizable networks with few anchors, *Proc. Algosensors,* 2006, (S. Nikoletseas and J. D. P. Rolim, eds.) Springer Lecture Notes in Computer Science 4240 (2006), pp. 176–183.

[10] J. Graver, B. Servatius and H. Servatius, *Combinatorial Rigidity,* AMS Graduate Studies in Mathematics Vol. 2 (1993).

[11] B. Hendrickson, Conditions for unique graph realizations, *SIAM J. Comput.,* **21** (1992), no. 1, 65–84.

[12] J. E. Hopcroft and R. E. Tarjan, Dividing a graph into triconnected components, *SIAM J. Comput.,* **2** (1973), 135–158.

[13] B. Jackson and T. Jordán, Connected rigidity matroids and unique realizations of graphs, *J. Combinatorial Theory Ser B,* Vol. **94** (2005), 1–29.

[14] B. Jackson and T. Jordán, The d-dimensional rigidity matroid of sparse graphs, *J. Combinatorial Theory Ser B,* Vol. **95** (2005), 118–133.

[15] B. Jackson and T. Jordán, Graph theoretic techniques in the analysis of uniquely localizable sensor networks, in: *Localization algorithms and strategies for wireless sensor networks,* G. Mao, B. Fidan (eds.), IGI Global (2009), pp. 146–173.

[16] B. Jackson, T. Jordán, and Z. Szabadka, Globally linked pairs of vertices in equivalent realizations of graphs, *Discrete and Computational Geometry,* Vol. **35** (2006), 493–512.

[17] N. Katoh and S. Tanigawa, On the infinitesimal rigidity of bar-and-slider frameworks, preprint (2009).

[18] G. Laman, On graphs and rigidity of plane skeletal structures, *J. Engineering Math.,* **4** (1970), 331–340.

[19] A. Lee, I. Streinu and L. Theran, The slider-pinning problem, *Proceedings of the 19th Canadian Conference on Computational Geometry,* Ottawa, Ontario (August 2007).

[20] M. Lorea, Hypergraphes et matroides, *Cahiers Centre Etud. Rech. Oper.,* **17** (1975), pp. 289–291.

[21] L. Lovász, The matroid matching problem, in: *Algebraic methods in graph theory,* Vols. I and II, Colloquia Mathematica Societatis János Bolyai 25, North-Holland, Amsterdam (1981), pp. 495–517.

[22] L. Lovász, Matroid matching and some applications, *J. Combin. Theory Ser. B,* **28** (1980), no. 2, 208–236.

[23] L. Lovász and Y. Yemini, On generic rigidity in the plane, *SIAM J. Algebraic Discrete Methods,* **3** (1982), no. 1, 91–98.

[24] M. Makai and J. Szabó, The parity problem of polymatroids without double circuits, *Combinatorica,* **28 (6)** (2008), pp. 679–692.

[25] A. Mansfield, On the computational complexity of a rigidity problem, *IMA J. Appl. Math.,* **27** (1981), no. 4, 423–429.

[26] A. Recski, *Matroid theory and its applications in electric network theory and in statics,* Akadémiai Kiadó, Budapest (1989).

[27] A. Schrijver, *Combinatorial optimization,* Springer (2003).

[28] B. Servatius, O. Shai and W. Whiteley, Combinatorial characterization of the Assur graphs from engineering, *Europ. J. Combin.,* in press.

[29] A. M. So and Y. Ye, Theory of semidefinite programming for sensor network localization, *Math. Program.,* **109** (2007), no. 2–3, Ser. B, 367–384.

[30] J. Szabó, Matroid parity and jump systems: a solution to a conjecture of Recski, *SIAM J. Discrete Math.,* Vol. 22, No. 3 (2008), pp. 854–860.

[31] W. Whiteley, Some matroids from discrete applied geometry. Matroid theory (Seattle, WA, 1995), 171–311, Contemp. Math., 197, Amer. Math. Soc., Providence, RI (1996).

[32] W. Whiteley, Rigidity and scene analysis, in: *Handbook of Discrete and Computational Geometry* (J. E. Goodman and J. O'Rourke, eds.), CRC Press, Second Edition (2004), pp. 1327–1354.

Tibor Jordán

Department of Operations Research,
Eötvös University,
Pázmány Péter sétány 1/C,
1117 Budapest,
Hungary

e-mail: jordan@cs.elte.hu

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 173–212.

Noise Sensitivity and Chaos in Social Choice Theory

GIL KALAI*

In this paper we study the social preferences obtained from monotone neutral social welfare functions for random individual preferences. It turns out that there are two extreme types of behavior. On one side, there are social welfare functions, such as the majority rule, that lead to stochastic stability of the outcome in terms of perturbations of individual preferences.

We identify and study a class of social welfare functions that demonstrate an extremely different type of behavior which is completely chaotic: they lead to a uniform probability distribution on all possible social preference relations and, for every $\varepsilon > 0$, if a small fraction ε of individuals change their preferences (randomly) the correlation between the resulting social preferences and the original ones tends to zero as the number of individuals in the society increases. This class includes natural multi-level majority rules.

1. Introduction

How likely is it that small random mistakes in counting the votes in an election between two candidates will reverse the election's outcome? And if there are three alternatives and the society prefers alternative a to alternative b and alternative b to alternative c how likely is it that a will be preferred to c? We will show that for general social welfare functions these two questions are closely related.

In this paper we consider the behavior of general social welfare functions with respect to random uniform voter profiles. Namely, the individual pref-

*I wish to thank Itai Benjamini, Guy Kindler, Nati Linial, Elchanan Mossel, Motty Perry, Oded Schramm, and Yuval Peres for fruitful discussions. This research was supported by an ISF Bikura grant, by an NSF grant, and by Microsoft, Israel.

erences on a set of alternatives are uniformly and independently distributed among all order relations. It turns out that there are two extreme types of behavior: social welfare functions, such as the majority rule, that lead to stochastic stability of the outcome in terms of perturbations of individual preferences and social welfare functions that lead to what we refer to as "social chaos."[1]

These two types of behavior in the case of two alternatives have been studied by Benjamini, Kalai and Schramm (1999).

A simple game (or voting game) G defined on a set N of players (voters) is described by a function v (called the "payoff function",) that assigns to every subset (coalition) S of players the value "1" or "0". We assume that $v(\emptyset) = 0$ and $v(N) = 1$. A candidate is elected if the set S of voters that voted for him is a winning coalition in G, i.e., if $v(S) = 1$. We will always assume that the game G is monotone, i.e., that if $v(R) = 1$ and $R \subset S$ then $v(S) = 1$. Recall that a simple game is *proper* if $v(S) + v(N \backslash S) \leq 1$ for every coalition S, i.e., if the complement of a winning coalition is a losing one. A simple game G is *strong* if $v(S) + v(N \backslash S) = 1$ for every coalition S, i.e., if it is proper and the complement of a losing coalition is a winning one.

It is convenient to let $N = \{1, 2, \ldots, n\}$ and regard v as a *Boolean function* $v(x_1, x_2, \ldots, x_n)$ where each variable x_i is a Boolean variable: $x_i \in \{0, 1\}$, and thus $v(x_1, x_2, \ldots, x_n)$ stands for $v(S)$, where $S = \{i \in N : x_i = 1\}$. This notation is used in much of the relevant mathematics and computer science literature. It is also consistent with the basic economic interpretation of the variables x_i as signals the voters receive regarding the superior alternative.

A neutral social welfare function $F(A)$ on a set A of m alternatives based on a strong simple game G is defined as follows: the function $F(A)$ is a map that assigns an asymmetric relation R on the alternatives to every profile of individual preferences. Given a monotone strong simple game G and a set A of alternatives, the society prefers alternative a to alternative b if the set of voters that prefers a to b forms a winning coalition.

[1] There are many studies of chaos and chaotic dynamics in economics but they appear to be rather different from this one. In the context of social choice theory, the huge number of "voting paradoxes" and the fact that "everything can happen" are often referred to as chaotic phenomena and were studied extensively by Saari; see, e.g., Saari (2001). Saari's works are concerned mainly with specific voting methods like majority, plurality, Borda's method, and some weighted versions of them. We refer to the property that "anything can happen", as "indeterminacy". Indeterminacy is different from yet related to our notion of social chaos; see Section 7 and Kalai (2004).

Remark. The term "neutral" refers to the property that the social welfare function is invariant under permutation of the alternatives. Neutrality requires that the simple games describing the choice between two alternatives be strong, and, in addition, that the same simple game is used for every pair of alternatives.[2]

We study probabilistic properties of social welfare functions for random uniform voter profiles. Thus, to every individual, we assign the same probability $(1/m!)$ for each one of the $m!$ order preference relations on the alternatives. Furthermore, the individual preference relations are independent. In other words, if there are n individuals, we assign the same probability for each one of the $(m!)^n$ possible profiles of individual preferences. This is a standard probabilistic model which has been extensively studied in the literature especially in the case of the majority rule (see Gehrlein (1997)). It is worth noting at this point that we will be studying the asymptotic properties of strong simple games and the associated social welfare functions, and it will be convenient to describe our results in terms of sequences $(G_k)_{k=1,2,...}$ of simple games.[3]

We can describe our notion of social chaos using several equivalent definitions and we will start with the following simple definition based on the probability of cyclic social preferences when there are three alternatives. Let G be a monotone strong simple game with n players and let F be a neutral social welfare function on three alternatives based on G. Consider a random uniform profile for the individual order preferences between the three alternatives. Let $p_{\mathbf{cyc}}(G)$ be the probability that the social preferences are cyclic. Arrow's theorem asserts that unless G is a dictatorship, $p_{\mathbf{cyc}}(G) > 0$.

Definition 1.1. The sequence (G_k) of monotone strong simple games leads to *social chaos* if

$$\lim_{k\to\infty} p_{\mathbf{cyc}}(G_k) = 1/4. \tag{1.1}$$

[2] The neutral social choice we consider satisfies the basic Pareto and IIA conditions. We do not assume that the social preference is necessarily an order relation. Sometimes they are referred to as "generalized social welfare functions."

[3] For our results we will not require any relation between the games G_k for different values of k. It is useful to think of the results applied to the case where the various G_k's represent the "same" type of voting rule, such as "simple majority" for different population sizes. But giving a formal definition of a "voting rule" that applies to populations of different sizes is not an easy matter and, in any case, is not needed in this paper.

There are altogether eight possible social asymmetric relations for three alternatives, six of which are order relations and two of which are cyclic. Consider a social welfare function based on a strong monotone simple game G on three alternatives and the uniform distribution on individual order preferences. Symmetry considerations imply that the probabilities of obtaining each of the six order relations as the social preference relation are equal, as are the two probabilities for the two cyclic relations. Therefore, another way of stating relation (1.1) is that for a social welfare function based on (G_k) on three alternatives under the uniform distribution on individual order preferences, for every asymmetric relation R, the probability, for a random voter profile, that the social preferences are described by R tends to $1/8$.

A result of Gulibaud (see Gehrlein (1997)) asserts that for the social welfare functions on three alternatives based on simple majority, the probability for a cyclic order relation as the number of voters tends to infinity is $1/4 - \big(3/(2 \cdot \pi)\big) \cdot \arcsin(1/3) \approx .08744$. In other words, if we assume that the voter profile is uniform and observe that the society prefers alternative a to b and b to c, our a posteriori probability of the society preferring a to c tends to 0.876. In contrast, in a similar scenario where the sequence G_k satisfies relation (1.1) our a posteriori probability of the society preferring a to c will tend to 0.5 (which is the a priori probability).[4]

We will now move from three alternatives to two and consider the effect of random noise. Suppose that there are two candidates and that the voters' preferences are random and uniform or equivalently that $x_i = 1$ with probability $1/2$ independently for all voters. *Noise sensitivity* is defined as the effect of random independent mistakes in counting the votes. Formally, for a strong simple game G and $t > 0$, consider the following scenario: first choose the voter signals $x_1, x_2, \ldots, x_n$ randomly such that $x_i = 1$ with probability $p = 1/2$ independently for $i = 1, 2, \ldots, n$. Let $S = v(x_1, x_2, \ldots, x_n)$. Next let $y_i = x_i$ with probability $1-t$ and $y_i = 1-x_i$ with probability t, independently for $i = 1, 2, \ldots, n$. Let $T = v(y_1, y_2, \ldots, y_n)$. Define $N_t(G)$ to be the probability that $S \neq T$.

Definition 1.2. A sequence $(G_k)_{k=1,2,\ldots}$ of strong simple games is *asymptotically noise-sensitive* if, for every $t > 0$,

$$\lim_{k \to \infty} N_t(G_k) = 1/2. \tag{1.2}$$

[4] (Kalai (2002) proved that for every monotone strong simple game this a posteriori probability is always larger than 0.5.)

Our first theorem shows a surprising relation between noise sensitivity for two alternatives and the probability of Condorcet's paradox for three alternatives.

Theorem 1.3. *A sequence (G_k) of monotone strong simple games leads to social chaos if and only if it is noise-sensitive.*

In proving Theorem 1.3 we obtain a simple formula which, for every strong simple game G, directly relates the probability for Condorcet's paradox and noise sensitivity when the probability for an error is $1/3$.

Proposition 1.4. *For every strong simple game G,*

$$2P_{\mathbf{cyc}}(G) + 1 = 3N_{1/3}(G). \tag{1.3}$$

We will now define the complementary notion of noise (or stochastic) stability.

Definition 1.5. A class $\mathcal{G}$ of strong simple games is *uniformly noise-stable* if for every $s > 0$ there is $u(s) > 0$ such that $G \in \mathcal{G}$,

$$N_{u(s)}(G) \leq s. \tag{1.4}$$

Theorem 1.6. *The family of simple majority functions M_n is uniformly noise-stable. Moreover, $N_{u(s)}(M_n) \leq s$ for $u(s) = K \cdot s^2$ for some constant K.*

Our next theorems provide several additional equivalent properties of social chaos which we believe justify this notion. We first note that we can consider slightly larger classes of games. A sequence G_k of monotone proper simple games has a *strongly diminishing bias* if for a random uniform voter profile, the probability of a "tie" tends to zero as n tends to infinity. An example is the class of simple majority games. All our results extend from strong simple games to the case of monotone proper simple games with strongly diminishing bias.

An asymmetric relation R on a finite set X is a binary relation such that every pair of elements $x, y \in X$ is ascribed one and only one of the relations xRy or yRx. Clearly, there are $2^{\binom{m}{2}}$ asymmetric relations on a set of m alternatives. For a monotone strong simple game G and an asymmetric relation R on a set of m alternatives, let $p_R(G)$ be the probability under the uniform distribution on voter profiles that the social welfare function based on G will lead to R as the social preference relation.

Theorem 1.7. *Let (G_k) be a sequence of proper strong simple games. The following properties P_m are equivalent for $m \geq 3$:*

(P_m): *For a social welfare function based on (G_k) on a set A of m alternatives and every asymmetric relation R on A,*

$$\lim_{k\to\infty} p_R(G_n) = 1/2^{\binom{m}{2}}. \tag{1.5}$$

Note that (P_3) is simply our definition of social chaos. It is easy to see that (P_m) implies $P_{m'}$ when $m' < m$. Surprisingly, the reverse implication also holds when $m' \geq 3$.

Consider a neutral social welfare function and a random voter profile, and let R be the resulting preference relation for the society. Consider the following scenario:

Suppose that every voter with a small probability t reconsiders and reverses the order of the two alternatives he ranked in the two *last* places. Let R' be the new social preference relation.

Theorem 1.8. *Let $m \geq 2$ be a fixed integer and t, $0 < t < 1/2$, be a fixed real number. A sequence G_n leads to social chaos if and only if any one of the following equivalent properties is satisfied:*

(A_m) *The correlation between R and R' tends to zero with n.*

(B_m) *For every two asymmetric relations R_1 and R_2 on A, the probability that $R' = R_2$, conditioned on $R = R_1$, tends to $1/2^{\binom{m}{2}}$ as n tends to infinity.*

Remark. Here we consider R and R' as probability distributions on asymmetric preference relations, and by talking about the correlation between R and R' we refer to the correlation between these two distributions. (Of course, B_m implies A_m, but as it turned out they are both equivalent to social chaos.)

(Properties A_2 and B_2 both coincide with the notion of noise sensitivity.)

In Section 2 we provide the mathematical tools and proofs for our results concerning the equivalence between the different properties of social chaos. We also describe there relations with the Banzhaf power index and additional characterization in terms of the correlation with weighted majority functions. Benjamini, Kalai and Schramm (1999) showed that the class of weighted majority (strong) games is uniformly noise-stable. A stronger quantitative version was proved by Peres (2004). Benjamini, Kalai

and Schramm (1999) also proved that if (G_k) is sequence of strong simple games that is asymptotically uncorrelated with weighted majority games, then it is noise-sensitive.

In Section 3 we will analyze several examples. Consider the class of multi-level majority-based voting rules described inductively as follows: (i) Simple majority on an odd number of at least three voters has one layer. (ii) Suppose that the players are divided into an odd number of three or more groups and on each group we consider a game with r layers. If G is defined by the simple majority of the outcomes, then G is said to be a multi-layer majority-based simple game with $r + 1$ layers.

Theorem 1.9. *A sequence (G_k) of multi-layer voting rules based on simple majority leads to social chaos if and only if the number of layers tends to infinity as k does.*

The reason for the chaotic behavior is that any new level of majority will amplify the probability of a cyclic outcome in three alternatives and by a similar argument every new level will amplify the noise introduced by mistakes in counting the votes.

When we base our rule on supermajority, social chaos may already appear for two layers. Consider a set of n individuals that is divided into many committees of size k. The decision between two alternatives is based on the majority of committees that prefer one of the alternatives by a 2/3-supermajority. This rule defines a proper simple game and in order to have a strongly diminishing bias we require that $k = o(\log n)$. We will show in Section 3 that this example leads to complete social chaos when the size of the committees tends to infinity with n. The reason for the chaotic behavior in this case is that the outcomes of the elections are determined by a small number of committees with "decisive views"; i.e., they are sufficiently biased towards one of the alternatives. Such a strong bias in favor of one of the alternatives is unlikely to survive following a small perturbation of the individual preferences. We note that hierarchical or recursive aggregation of preferences similar to those in the first example, and cases where the social preferences heavily depend on a small number of small communities with decisive views, can both represent realistic situations of preference aggregation.

The assumption of uniform and independent voter behavior is a standard one and is the basis of a large body of literature. Most of this literature deals with the majority voting rule (see, e.g., Bell (1981) and Gehrlein (1997)). Assuming uniform and independent voter behavior is unrealistic, though it

can be argued that for issues concerning noise sensitivity and Condorcet's paradox this assumption represents a realistic way of comparing different voting methods. In Sections 4 and 5 we examine our probabilistic assumptions and discuss the situation under more general probability distributions for voter signals/behavior. Section 4 is devoted to the case that either the probability distribution or the voting rule are a priori biased. In Section 5 we present an example of chaotic behavior under the majority voting rule, when voter behavior is not independent.

Perhaps the most important form of "noise" in real-life elections and other forms of aggregated choice is abstention. In Section 6 we extend our model to allow for individual indifference between alternatives. Analyzing this model, we find out that noise sensitivity implies that a small change in the fraction of voters who decide to abstain has a dramatic effect on the outcomes, similar to the effect of random mistakes in counting their votes. Section 7 is devoted to a discussion of issues related to noise stability, information aggregation, indeterminacy, the asymptotic nature of our results, and possible extensions to other economic models. Section 8 concludes.

2. Proofs of the Equivalence Theorems

In this section we prove the various equivalent properties of sequences of simple games that lead to social chaos. We also state additional equivalent formulations. The proofs rely on two methods. A rather simple coupling argument allows us to move from noise sensitivity to the conclusions concerning more than two alternatives given in Theorems 1.3, 1.7, and 1.8. The only remaining step is to show that our definition of social chaos in terms of the probability of Condorcet's paradox for three alternatives implies noise sensitivity. I am not aware of a direct probabilistic argument and the proof of this implication relies on a rather elementary harmonic analysis of Boolean functions. This part of the proof requires mathematical techniques and concepts developed in Benjamini, Kalai and Schramm (1999) and in Kalai (2002).

Let (G_k) be a sequence of strong simple games. (All our arguments apply unchanged to the case of proper simple games with strongly diminishing bias.) We consider the following properties:

- (NS) Noise sensitivity
- (P_m) Asymptotically uniform distribution of social preference relations when there are m alternatives, $m \geq 3$
- Properties (A_m), (B_m), $m \geq 2$.

We wish to prove that all these properties are equivalent.

2.1. A coupling argument

Proof of the implication $(NS) \Rightarrow (P_k)$**.** We rely here on a simple (coupling) idea which nonetheless requires careful application. Consider a random voter profile. For two alternatives a and b consider the set V of voters for whom alternatives a and b are consecutive. We will try to determine the effect of reversing the order of the two alternatives a and b with probability $1/2$ for all voters in V.

Let R and R' be two asymmetric relations that differ only for the two alternatives a and b, namely, for a set $\{c,d\} \neq \{a,b\}$ of two alternatives cRd if and only if $cR'd$ and for the two alternatives a and b, aRb and $bR'a$. We will derive from noise sensitivity that

$$\lim_{k\to\infty} p_R(G_k) = \lim_{k\to\infty} p_{R'}(G_k). \tag{2.1}$$

Consider the following process of determining two profiles P and P' of individual preference relations:

Step (1): Determine the preference order relations between alternatives a and b for every voter randomly and uniformly.

Step (2): Choose the order preference relations for all voters subject to the voters' preference between a and b that was determined in step (1).

Let P be the resulting profile of voters' order relations obtained by this process. Of course, P is a uniformly distributed random profile.

Let V be the set of voters for whom alternatives a and b are consecutive in the preference relation. The set V is a random subset of the set of voters. The probability that a voter i belongs to V is $1/(m-1)$ and these events are independent. Moreover, the set V and the random voter preference relations are independent.

Let P' be the voter profile obtained from P by reversing the order relations between alternatives a and b for every voter in V. Note that P' is also a uniformly distributed random profile.

Let R be the asymmetric relation obtained by applying the social welfare function F to the profile P and let R' be the asymmetric relation obtained by applying F to P'. Note that the two relations R and R' coincide for every pair of alternatives except possibly a and b.

If we focus on these two variables we note that the distribution of voter preference relations between a and b in P is random. We claim that the correlation between the events aRb and $aR'b$ tends to zero as n tends to infinity. (In other words, the probability that $aR'b$ given that aRb tends to 1/2.) Indeed, this follows from the fact that $\lim_{n\to\infty} N_t(G_n) = 1/2$ for $t = 1/(m-1)$. ■

The proof of the implication $(NS) \Rightarrow (B_m)$ is similar but a little more complicated to the proof presented above. We omit the details.

Proof of the implication $(P_m) \Rightarrow (P'_m)$ for $m' \leq 3$, $m' < m$. Let A be a set of m alternatives and let A' be a subset of A containing m' alternatives. A random uniform voter profile on A induces a random uniform voter profile on A'. Every asymmetric relation R' on A' can be extended to precisely $2^{\binom{m}{2}-\binom{m'}{2}}$ asymmetric relations R on A. Given such an asymmetric relation R, the probability that a social welfare function based on G_k will lead to R tends to $1/2^{\binom{m}{2}}$. (By our assumption (P_m).) Therefore the probability that the restriction on A' will lead to R' tends to $1/2^{\binom{m'}{2}}$ as required. ■

The same argument shows the implications: $(A_m) \Rightarrow (A_{m'})$, for $2 \leq m' < m$.

2.2. An elementary harmonic analysis argument

Proof of Theorem 1.3. As mentioned in the Introduction, a simple game can be described by a *Boolean function*, namely, $f(x_1, x_2, \ldots, x_n)$, where the variables x_k take the values 0 or 1 and the value of f itself is also either 0 or 1. Every 0-1 vector $x = (x_1, x_2, \ldots, x_n)$ corresponds to a subset of players $S = \{k : x_k = 1\}$ and we let $f(x_1, x_2, \ldots, x_2) = v(S)$. In other words, the

Boolean function $f(x_1, x_2, \ldots, x_n)$ is just a slightly different notation for the payoff function $v(S)$, and we will refer also to f as the payoff function of the game.

Let Ω_n denote the set of 0-1 vectors $(x_1, \ldots, x_n)$ of length n. Let $L_2(\Omega_n)$ denote the space of real functions on Ω_n, endowed with the inner product

$$\langle f, g \rangle = \sum_{(x_1, x_2, \ldots, x_n) \in \Omega_n} 2^{-n} f(x_1, \ldots, x_n) g(x_1, \ldots, x_n).$$

The inner product space $L_2(\Omega_n)$ is 2^n-dimensional. The L_2-norm of f is defined by

$$\|f\|_2^2 = \langle f, f \rangle = \sum_{(x_1, x_2, \ldots, x_n) \in \Omega_n} 2^{-n} f^2(x_1, x_2, \ldots, x_n).$$

Note that if f is a Boolean function, then $f^2(x)$ is either 0 or 1 and therefore $\|f\|_2^2 = \sum_{(x_1, \ldots, x_n) \in \Omega_n} 2^{-n} f^2(x)$ is simply the probability that $f = 1$ (with respect to the uniform probability distribution on Ω_n). If the Boolean function f represents a strong simple game then $\|f\|_2^2 = 1/2$.

For a subset S of N consider the function

$$u_S(x_1, x_2, \ldots, x_n) = (-1)^{\sum \{x_i : i \in S\}}.$$

It is not difficult to verify that the 2^n functions u_S for all subsets S form an orthonormal basis for the space of real functions on Ω_n.

For a function $f \in L_2(\Omega_n)$, let

$$\widehat{f}(S) = \langle f, u_S \rangle$$

($\widehat{f}(S)$ is called a Fourier–Walsh coefficient of f). Since the functions u_S form an orthogonal basis it follows that

$$\langle f, g \rangle = \sum_{S \subset N} \widehat{f}(S) \widehat{g}(S). \tag{2.2}$$

In particular,

$$\|f\|_2^2 = \sum_{S \subset N} \widehat{f}^2(S). \tag{2.3}$$

This last relation is called Parseval's formula.

Remark. We will demonstrate now the notions introduced here with a simple example. Let f_3 denote the payoff function of the simple majority game with three players. Thus, $f_3(x_1, x_2, x_3) = 1$ if $x_1 + x_2 + x_3 \geq 2$ and $f(x_1, x_2, x_3) = 0$, otherwise. The Fourier coefficients of f_3 are easy to compute: $\widehat{f_3}(\emptyset) = \sum(1/8)f_3(x) = 1/2$. In general if f is a Boolean function then $\widehat{f}(\emptyset)$ is the probability that $f(x) = 1$ and when f represents the payoff function of a strong simple game $\widehat{f}(\emptyset) = 1/2$. Next, $\widehat{f_3}\big(\{1\}\big) = 1/8\big(f_3(0,1,1) - f_3(1,0,1) - f_3(1,1,0) - f_3(1,1,1)\big) = (1-3)/8$ and thus $\widehat{f_3}\big(\{j\}\big) = -1/4$, for $j = 1, 2, 3$. Next, $\widehat{f_3}(S) = 0$ when $|S| = 2$ and finally $\widehat{f_3}\big(\{1,2,3\}\big) = 1/8\big(f_3(1,1,0) + f_3(1,0,1) + f_3(0,1,1) - f(1,1,1)\big) = 1/4$.

We point out the following simple result:

Proposition 2.1. *If a Boolean function f is the payoff function of a strong simple game then $\widehat{f}(S) = 0$ whenever $|S|$ is an even positive integer.*

Proof. Let $g = f - 1/2$ $(= f(x) - 1/2u_\emptyset)$. Since f represents a strong simple game $g(1 - x_1, 1 - x_2, \ldots, 1 - x_n) = -g(x_1, x_2, \ldots, x_n)$. When $|S|$ is even, consider the contributions of $(x_1, x_2, \ldots, x_n)$ and $(1 - x_1, 1 - x_2, \ldots, 1 - x_n)$ to the expression $\widehat{g}(S) = \langle g, u_S \rangle$. Note that these two contributions cancel out and therefore $\widehat{g}(S) = 0$ for every set S of even size. It follows that when S is a nonempty set of even size, $\widehat{f}(S) = \widehat{g}(S) = 0$. ■

Let f be a Boolean function. For $k \geq 0$, let

$$W_k(f) = \sum \big\{ \widehat{f}^2(S) \,:\, S \subset N,\ |S| = k \big\}. \tag{2.4}$$

Theorem 2.2. *The probability $P_{\mathbf{cyc}}(G)$ of a cyclic outcome of a social welfare function on three alternatives based on a strong simple game G with a payoff function f is*

$$P_{\mathbf{cyc}}(G) = 1/4 - \sum_{k=1}^{n} (1/3)^{k-1} W_k(f). \tag{2.5}$$

Theorem 2.2 is from Kalai (2002).

Theorem 2.3. *Let G be a strong simple game and let f be its payoff function. The probability that an ε-noise will change the outcome of an election is given by the formula:*

$$N_\varepsilon(G) = 1/2 - 2 \cdot \sum_{k=1}^{n} (1 - 2\varepsilon)^k W_k(f). \tag{2.6}$$

Theorem 2.3 is from Benjamini, Kalai and Schramm (1999). The proofs of Theorems 2.2 and 2.3 are quite elementary and rely essentially on relations (2.2) and (2.3). Note that we obtain the simple relation (Proposition 1.4) between the probability for Condorcet's paradox and noise sensitivity when the probability for a noisy bit is 1/3.

$$2P_{\mathbf{cyc}}(G) + 1 = 3N_{1/3}(G). \tag{2.7}$$

From Theorem 2.3 we can easily derive (see Benjamini, Kalai and Schramm (1999)):

Corollary 2.4. *A sequence (f_n) of Boolean functions that represent strong simple games is noise-sensitive if and only if for every $k > 0$*

$$\lim_{n\to\infty} \sum_{i=1}^{k} W_i(f_n) = 0. \tag{2.8}$$

As a matter of fact Corollary 2.4 holds as stated when the Boolean functions (f_n) do not necessarily represent strong simple games but rather satisfy that the sequence of probabilities that $f_n = 1$ is bounded away from 0 and 1 as n tends to infinity.

Proof of Theorem 1.3. For this proof we need to compare the information given by Theorem 2.2 and Corollary 2.4. Suppose that $P_{\mathbf{cyc}}(G_n) \to 1/4$. Let f_n be the payoff functions of the strong simple game G_n. It follows from relation (2.5) that for every $r > 0$, $\lim_{n\to\infty} W_r(f_n) = 0$, which by Corollary 2.4 is equivalent to noise sensitivity. On the other hand, if (f_n) satisfies relation (2.8) then it follows from relation (2.5) that $\lim_{n\to\infty} P_{\mathbf{cyc}}(G_n) = 1/4$. (This implication was also proved earlier.) ■

We now complete the proof of Theorems 1.7 and 1.8. We have already proved the following implications: $(P_3) \Rightarrow (NS) \Rightarrow (P_m) \Rightarrow (P_3)$ and $(NS) \Rightarrow (B_m)$, for every $m \geq 3$. Of course, if property (P_m) holds then property (A_m) is equivalent to property (B_m) and therefore $(NS) \Rightarrow (B_m)$, $m \geq 3$. On the other hand, $(B_m) \Rightarrow (B_2) = (NS)$ and $(A_m) \Rightarrow (A_2) = (NS)$.

Remark. We now continue to demonstrate the notions and results introduced here with our simple example. As before, f_3 denote the payoff function of the simple majority game with three players. As we saw, the

Fourier coefficients of f_3 are described by $\widehat{f_3}(\emptyset) = 1/2$, $\widehat{f_3}(\{j\}) = -1/4$, for $j = 1, 2, 3$, $\widehat{f_3}(S) = 0$ when $|S| = 2$ and $\widehat{f_3}(\{1,2,3\}) = 1/4$. When there are three alternatives the probability for cyclic social preferences given by formula (2.5) is $1/4 - 3 \cdot (1/4)^2 - (1/9) \cdot (1/4)^2 = 8/(9 \cdot 16) = 1/18$. This value agrees with the well-known outcome given by a direct computation. The value of $N_t(f_3)$ is $1/2 - 2 \cdot (1-2t)(3/16) - 2 \cdot (1-2t)^3 \cdot (1/16) = 3t/2 - 3/2t^2 + t^3$. (Again, this can be derived by a direct computation.)

2.3. Individual power

We conclude this section with a short discussion of the relation between noise sensitivity and the Banzhaf power index. Let $b_j(G)$ denote the Banzhaf power index of player j in a strong simple game G. Recall that $b_j(G)$ is defined as the probability that player j is pivotal, i.e., the probability that for a random coalition S not containing j it is the case that S is a losing coalition but $S \cup \{j\}$ is a winning one. Let $b_{\max}(G)$ denote the maximum value of the Banzhaf power indices and let $I(G)$ denote the sum of the Banzhaf power indices for all players in a simple game, known also as the total influence of the game. It is known that in the case of n players, $I(G)$ is maximized by a simple majority game. (See, e.g., Friedgut and Kalai (1996).) For simple majority games $I(G)$ is proportional to $\sqrt{n}$.

Definition 2.5. The sequence (G_k) of monotone simple games has a *diminishing individual Banzaf power* if

$$\lim_{k \to \infty} b_{\max}(G_k) = 0.$$

Proposition 2.6. *If the sequence (G_k) of monotone simple games leads to social chaos then it has a diminishing individual Banzhaf power.*

The proof is immediate: when there is a player with a substantial Banzhaf power the outcome of the game has a substantial correlation with this player vote. A small noise in counting the votes will miss the vote of this player with a substantial probability and therefore the correlation of the outcomes before and after the noise will also be substantial. (Here "substantial" means bounded away from zero.)

Definition 2.7. The sequence (G_k) of monotone simple games has a *bounded power ratio* if the ratio $b_{\max}(G_k)/b_{\min}(G_k)$ is uniformly bounded.

Given a monotone simple game G, let $p(G)$ be the probability that a random uniform subset of players is a winning coalition. If G is strong, then $p(G) = 1/2$. Given two monotone simple games G and H on the same set N of players let $p(G, H)$ be the probability that a random uniform subset S of players is a winning coalition for *both* G and H. The correlation between G and H, denoted by $\mathrm{cor}\,(G, H)$, is defined by

$$\mathrm{cor}\,(G, H) = p(G, H) - p(G)p(H).$$

The well-known FKG inequality asserts that for every two monotone simple games $\mathrm{cor}\,(G, H) \geq 0$.

Definition 2.8. Two sequences (G_k) and (H_k) of monotone strong simple games are *asymptotically uncorrelated* if

$$\lim_{n\to\infty} \mathrm{cor}\,(G_n, H_n) = 0.$$

Weighted majority games are important class of simple games. Given n nonnegative weights $w_1, w_2, \ldots, w_n$, and a value W, define the payoff function $v(S)$ to be one if $\sum_{i\in S} w_i > W$ and $v(S) = 0$, otherwise. (If $W = (1/2)\sum_{i=1}^{n} w_i$ and no partial sum of weights is equal to W the weighted majority game is strong.)

Theorem 2.9. *A sequence (G_k) of monotone strong simple games leads to social chaos if and only if it is asymptotically uncorrelated with every sequence of weighted majority games.*

An important special case is:

Theorem 2.10. *A sequence (G_k) with a bounded Banzhaf power ratio leads to social chaos if and only if it is asymptotically uncorrelated with the sequence of simple majority games.*

Theorem 2.10 together with our earlier results demonstrates the "robustness of majority".[5] Voting methods (based on proper simple games where the voters have comparable power and the probability of a tie tends to zero) either have a substantial correlation with simple majority or else lead to social chaos.

Theorem 2.11. *Let (G_n) be a sequence of monotone strong simple games with a bounded power ratio and suppose that G_n has n players. The sequence G_n leads to social chaos if and only if $\lim_{n\to\infty} I(G_n)/\sqrt{n} = 0$.*

[5] A term coined by Dasgupta and Maskin in another context.

This result can be derived directly from the following result from Benjamini, Kalai and Schramm (1999):

Theorem 2.12. *The sequence (G_k) is noise-sensitive if and only if*

$$\lim_{k\to\infty} \sum \big(b_j(G_k)\big)^2 = 0.$$

(Here the sum extends over all players of the game G_k.)

Remark. Holzman, Lehrer and Linial (1988) proved that for every monotone simple game $\sum \big(b_k(G)\big)^2 \leq 1$.

3. Multi-Level Majority

3.1. Simple majority

Our principal examples of social welfare functions that exhibit social chaos are based on multi-level majority. For their analysis we will have to start with the case of simple majority. The outcomes of simple majority under random individual voter profiles have been the subject of intense study. We will first require a few results concerning the behavior of simple majority. Given a monotone strong simple game G recall that $p_{\mathbf{cyc}}(G)$ denotes the probability that for a social welfare function on three alternatives A, B and C and a random uniform voter profile the social preferences will be cyclic.

We start with a simple result concerning the majority voting rule with an odd number of voters.

Lemma 3.1. *Let M_n be a simple majority game on an odd number n of players, $n \geq 3$. Then,*

(i) $N_\varepsilon(M_n) \geq N_\varepsilon(M_3) = 3/2 \cdot \varepsilon - 3/2\varepsilon^2 + \varepsilon^3$;

(ii) $p_{\mathbf{cyc}}(M_n) \geq p_{\mathbf{cyc}}(M_3) = 1/18$.

The proof is given in Section 9.1 in the Appendix. We point out that Fishburn, Gehrlein and Maskin proved that $p_{\mathbf{cyc}}(M_{2k+1}) \geq p_{\mathbf{cyc}}(M_{2k-1})$ for every $k \geq 3$. This implies, in particular, part (ii) of Lemma 3.1. The next result describes the limiting behavior of majority when the number of voters tends to infinity.

Proposition 3.2. (1) (*Sheppard* (1899)) *For every* $\varepsilon > 0$

$$\lim_{n\to\infty} N_\varepsilon(M_n) = \frac{\arccos(1-2\varepsilon)}{\pi}. \tag{3.1}$$

(2) *Gulibaud* (*see Gehrlein* (1997))

$$\lim_{n\to\infty} p_{\mathbf{cyc}}(M_n) = 1/4 - \big(3/(2\cdot\pi)\big)\cdot \arcsin(1/3) \approx .08744.$$

Proposition 1.4 gives a direct link between Gulibaud's formula and Sheppard's formula. Relation (2.7) asserts that for every n, $p_{\mathbf{cyc}}(M_n) = 3/2N_{1/3}(M_n) - 1/2$. Together with the case $\varepsilon = 1/3$ of Sheppard's formula $\lim_{n\to\infty} N_{1/3}(M_n) = \frac{\arccos(1/3)}{\pi}$ we obtain the formula of Gulibaud.

It follows from relation (3.1) that as ε tends to zero $\lim_{n\to\infty} N_\varepsilon(M_n) = \big(1+o(1)\big)\cdot\sqrt{2\varepsilon}/\pi$. This is a more precise version of Theorem 1.6. We will not reproduce here the proofs of Proposition 3.2 but rather will explain informally why, for simple majority on a large number n of players and for a small amount ε of noise, the function $N_\varepsilon(M_n)$ behaves like $\sqrt{\varepsilon}$. The median gap between the number of votes of the winning candidate and the losing one in a simple majority is close to $\sqrt{n}$. According to the central limit theorem the probability that the gap is larger than $t\sqrt{n}$ behaves like $e^{-t^2/4}$. It follows that a large gap is extremely rare while below the median gap the distribution of the gap is rather close to uniform. When we flip εn random votes the median gap between the number of votes for Alice that now go to Bob and votes for Bob that now go to Alice is thus close to $\sqrt{\varepsilon}\sqrt{n}$, and again it is rare that this gap is much larger. In order for the noise to change the election's results we require (apart from an event with a small probability) that the original gap between the votes for Alice and Bob be around $\sqrt{\varepsilon}\sqrt{n}$ and this occurs with probability of roughly $\sqrt{\varepsilon}$. On the other hand, if the gap between votes for Alice and votes for Bob behaves like $\sqrt{\varepsilon}\sqrt{n}$ then the probability that the noise will change the election's outcome is bounded away from zero. This argument indeed shows that $N_\varepsilon(M_n)$ behaves like $\sqrt{\varepsilon}$.

3.2. Multi-level games based on simple majority

The first class of examples that exhibit social chaos are based on multi-level majority where the number of levels increases. Recall the class of multi-level majority-based voting rules described inductively in the Introduction: (i) Simple majority on an odd number of at least three have one layer.

(ii) Suppose that the players are divided into an odd number of three or more groups and on each group we consider a game with r layers. If G is defined by the simple majority of the outcomes then G is said to be a multi-layer majority-based simple game with $r+1$ layers.

Define $u^{(1)}(\varepsilon) = u(\varepsilon) = 3/2 \cdot \varepsilon - 2\varepsilon^2 + 5/4\varepsilon^3$ and $u^{(r)}(\varepsilon) = u\big(u^{(r-1)}(\varepsilon)\big)$, $r = 2, 3, 4, \ldots$.

Theorem 3.3. *Let G_r be an r-layer simple game based on simple majority. Then*

$$N_\varepsilon(G_r) \geq u^{(r)}(\varepsilon).$$

As ε tends to 0,

$$N_\varepsilon(G_r) \geq (3/2)^r \varepsilon + O(\varepsilon^2).$$

Proof. The proof is by induction on r and the case $r = 1$ is given by Lemma 3.1. Consider a random vector $(x_1, x_2, \ldots, x_n) \in \Omega_n$. The value of $v(x_1, x_2, \ldots, x_n)$ is the simple majority of the $(r-1)$ level outcomes $z_1, z_2, \ldots, z_m$. Suppose that we flip each variable x_i with probability ε. By the induction hypothesis each z_i is flipped with probability of at least $u^{(r-1)}(\varepsilon)$ and by Lemma 3.1 the probability that v is flipped is at least $u\big(u^{(r-1)}\big)(\varepsilon) = u^{(r)}(\varepsilon)$. When $G = T_r$ is the recursive ternary majority game with r levels we obtain that $N_\varepsilon(T_r) = u^{(r)}(\varepsilon)$. To prove Theorem 3.3 we need to verify that for every $\varepsilon > 0$, $\lim_{r\to\infty} u^{(r)}(\varepsilon) = 1/2$.

We will state separately the result for a bounded number of levels.

Theorem 3.4. *Let $\phi(\varepsilon) = \phi^{(1)}(\varepsilon) = \frac{\arccos(1-\varepsilon)}{\pi}$ and let $\phi^{(i+1)}(\varepsilon) = \phi(\phi^{(i)}(\varepsilon)$, $i = 1, 2, \ldots$. Let r be a fixed positive integer. Consider a sequence (G_n) of r-level majority strong simple games where each level is based on simple majority of t_n players. (Thus, $n = t_n^r$.) Then for a fixed $\varepsilon > 0$ as n tends to infinity,*

$$N_\varepsilon(r) =: \lim_{n\to\infty} N_\varepsilon(G_n) = \phi^{(r)}(\varepsilon).$$

As ε tends to zero

$$N_\varepsilon(r) \geq \big(1 + o(1)\big)\,(2/\pi)\varepsilon^{1/2^r}.$$

In our model, simple majority is considerably more stable in the presence of noise than a two-layer election method like the U.S. electoral system. In an electoral system with many states and many voters in each state, $N_\varepsilon(G)$ is proportional, for small values of ε, to $\varepsilon^{1/4}$. In a three-stage system of this kind $N_\varepsilon(G)$ behaves like $\varepsilon^{1/8}$. Is the advantage of simple majority

on a two-level majority in terms of noise sensitivity relevant for evaluating the U.S. electoral method? Can we conclude that events like those in the 2000 U.S. presidential election are less likely to occur in a simple majority system? While the probabilistic models are unrealistic, it appears that the advantage of simple majority compared to a two-level electoral method is quite general.

Remark. The hierarchical voting methods considered here resemble the multi-tier system of councils ("soviets" in Russian). Lenin (among others) advocated this system during the 1917 Russian Revolution. Lenin's concept of centralized democracy is based on a hierarchical method of voting and was implemented in the Soviet Union and its satellites for party institutions, national bodies, and labor unions. (For national bodies, the method was changed in 1936.) For party institutions (which were the most important) there could be as many as seven levels. Party members in a local organization, for example, the Department of Mathematics in Budapest, elected representatives to the Science Faculty party committee who in turn elected representatives to the University council. The next levels were the council of the 5th District of Budapest, the Budapest council, the Party Congress, the Central Committee and finally the Politburo.[6] Friedgut (1979) is a good source on the early writings of Marx, Lenin, and others, and for an analysis of the Soviet election systems that were prevalent in the 70's.

3.3. Two-level games based on supermajority

The second class of examples of social welfare functions that exhibit social chaos are based on a two-level method where the lower level is biased. Consider a two-level proper game $G = G[a, b, t]$, where t satisfies $b/2 < t < b$, defined as follows. There are $a \cdot b$ voters divided into a communities of b voters each. Given a subset S of voters, we call a community C positive if $|S \cap C| \geq t$ and negative if $|S \cap C| \leq b - t$. The subset S is a winning coalition if there are more positive communities than negative communities. Let $G_n = G[a_n, b_n, t_n]$ be a sequence of such games. In order for G_n to have strongly diminishing bias it is necessary and sufficient that the expected number of decisive coalitions tends to infinity.

[6] I am thankful to Laci for this example.

Theorem 3.5. *If $G_n = G[a_n, b_n, t_n]$ has strongly diminishing bias and*

$$\lim_{n\to\infty} b_n = \infty,$$

then the sequence G_n leads to social chaos.

The proof is based on a straightforward but tedious computation and is omitted.

Remark. It can be shown by similar proofs to those we present that social chaos occurs for more general hierarchical voting methods. This applies both in the case where in each level the method is balanced and the number of levels tends to infinity and in the case where there are as little as two levels and the method is biased in the lower level. Social chaos may occur also for more complicated recursive methods. For example the different communities that we considered need not be disjoint. But the analysis in such cases can be considerably harder.

4. Bias

The framework described in this paper extends to more general probabilistic distributions of voter behavior (or individual signals). We expect that the dichotomy described here between stochastically stable and chaotic behavior can also be extended, though such extensions are conceptually and mathematically more difficult. When we consider the effect of stochastic perturbations there are also various possible models for the "noise." Does it represent mistakes in counting the votes? or, perhaps, a small fraction of voters reconsidering their positions or considering whether or not to vote? We also expect that the stochastic stability of the majority rule extends to various other probability distributions representing the voters' behavior.

Our probabilistic model assumes three properties which we will discuss separately:

(1) Independence: Voters' probabilities (or signals) are independent.

(2) No bias: For two alternatives a and b the probability that the society prefers alternative a to alternative b is $1/2$.

(3) Identical voters: Voters' probabilities (or signals) are identical.

The assumption of identical voter preference distribution is unrealistic but I expect that the basic results of this paper can be extended to cases where the assumption of identical voter behavior is relaxed while the other two assumptions are kept. We will concentrate on the other two assumptions and discuss in this section bias and in the next section dependence between voters.

The assumption of no bias is clearly an ideal assumption. A more realistic assumption in many cases of aggregated choice is the assumption of "small bias," which is defined as follows:

Definition 4.1. Consider a sequence (G_k, ν_k), where G_k is a proper simple game with n_k players and ν_k is a probability distribution on coalitions of G_k. (Alternatively, ν_k can be regarded as a probability distribution on 0–1 vectors of length n_k.) The sequence (G_k, ν_k) has *small bias* if for some constant $t > 0$, the probability $p_{\nu_k}(G_k)$ that a random coalition according to ν_k is a winning coalition satisfies

$$t \leq p_{\nu_k}(G_k) \leq 1 - t, \quad \text{for every} \quad k \geq 1.$$

While it is hard to justify that a priori two candidates in an election have the same probability of winning it is often quite realistic that the a priori probability of winning is substantial for both. We will return to this issue towards the end of this section.

We will describe in this section the extension of our results to cases with bias. Bias occurs when the simple game is strong and thus neutral between the candidates but the probability distribution of signals favors one of the candidates, and also when the probability distribution is uniform but the voting method itself is described by a simple game giving advantage to one of the candidates. The results in these two types of extensions (and their combination) are similar. The dichotomy between noise-stable and noise-sensitive monotone simple games and many of the results presented in this paper can be extended to the case of independent voter behavior with a small bias.

Recall that Ω_n denote the set of $0-1$ vectors $(x_1, x_2, \ldots, x_n)$ of length n. Given a real number p, $0 < p < 1$, $\mathbf{P}_p$ will denote that product distribution on Ω_n defined by

$$\mathbf{P}_p(x_1, x_2, \ldots, x_n) = p^k(1-p)^{n-k},$$

where $k = x_1 + x_2 + \cdots + x_n$. For a monotone simple game G with payoff function $v = v(x_1, x_2, \ldots, x_n)$, define

$$\mathbf{P}_p(G) = \sum \left\{ \mathbf{P}_p(x) v(x) \, : \, x \in \Omega_n \right\}.$$

Let $t > 0$ and consider the following scenario: first choose the voters' signals $x_1, x_2, \ldots, x_n$ at random such that $x_i = 1$ with probability p independently for $i = 1, 2, \ldots, n$. Consider $S = v(x_1, x_2, \ldots, x_n)$. Next let $y_i = x_i$ with probability $1 - t$ and $y_i = 1 - x_i$ with probability t, independently for $i = 1, 2, \ldots, n$. Let $T = v(y_1, y_2, \ldots, y_n)$. Let $C_t^p(G)$ be the correlation between S and R.

Definition 4.2. Consider a sequence $(G_k, p_k)_{k=1,2,\ldots}$ where G_k is a monotone simple game and $0 < p_k < 1$. The sequence (G_k, p_k) is *asymptotically noise-sensitive,* if for every $t > 0$,

$$\lim_{k \to \infty} C_t^{p_k}(G_k) = 0.$$

In the case of small bias, namely, when $\mathbf{P}_{p_k}(G_k)$ is bounded away from 0 and 1, Theorem 1.7, which asserts that simple majority is noise-stable, extends and so do Theorems 2.9 and 2.12.

If $\lim_{k \to \infty} \mathbf{P}_{p_k}(G_k) = 1$ (or 0) the sequence is *always* noise-sensitive.

Theorem 4.3. *Every sequence (G_k, p_k) of monotone simple games where $\lim_{n \to \infty} \mathbf{P}_{p_k}(G_n) = 1$ (or 0) is noise-sensitive.*

Theorem 4.3 follows from a Fourier-theoretic interpretation of noise sensitivity, which extends the formula for the uniform distribution, and a certain inequality of Bonami and Beckner.

Perhaps some explanation of the meaning of noise sensitivity in the presence of bias is in order. When there are two alternatives noise sensitivity can be described in two ways:

(i) The result following a small perturbation is asymptotically uncorrelated with the original result.

(ii) The probability that a small perturbation will change the collective choice tends to 1/2.

The reason that when there is no bias (i) implies (ii) is that the model assigns equal probabilities to the two outcomes. Property (i) may extend to rather general probability distributions but the dramatic conclusion in (ii) requires that there be no bias.

Assuming small bias, condition (i) implies the following:

(iii) With a probability uniformly bounded away from zero, a small amount of noise will change the outcome of the election.

Note that for a sequence of games (G_k) with small bias in order to conclude property (iii) it is enough to assume a weaker property than noise sensitivity. Property (iii) follows if the sequence (G_k), in addition to having small bias, is *uniformly chaotic,* which is defined as follows:

Definition 4.4. Consider a sequence $(G_k, p_k)_{k=1,2,...}$ where G_k is a monotone proper simple game and $0 < p_k < 1$. The sequence (G_k, p_k) is *uniformly chaotic,* if there is $\alpha > 0$ so that for every $t > 0$,

$$\lim_{k\to\infty} C_t^{p_k}(G_k) < 1 - \alpha.$$

It is interesting to compare what happens to conditions (i) and (iii) when the bias get larger (maintaining the assumption of independent and identical voter preferences). In an election between Alice and Bob, when the probability of Alice winning tends to zero then so does the probability that a small random perturbation in voter behavior will change the outcome of the election. On the other hand, Theorem 4.3 asserts that in this case it is *always* true, even for the majority rule, that the correlation between outcomes before and after the noise also tends to zero. Roughly speaking, Theorem 4.3 asserts that when your probability of winning the election is small and you have the option of forcing a recount of a random fixed fraction of votes, this option almost doubles your chances of winning (for the majority rule and for any other voting rule).

Our next proposition, whose proof follows directly from the definition of uniform noise stability, asserts that the property of noise stability is itself stable under small changes in the probability distribution.

Let $n(k)$ be a monotone sequence of positive integers. Consider two sequences (ν_k) and (ν'_k) where ν_k and ν'_k are probability distributions on $\Omega_{n(k)}$. We say that ν and ν' are asymptotically non-singular if for every sequence of events $S_k \in \Omega_{n(k)}$, $\lim \nu(S_k) = 0$ if and only if $\lim \nu'(S_k) = 0$.

Proposition 4.5. *Consider a family $\mathcal{G}$ of strong simple games which is uniformly noise-stable. Suppose that p_n is a sequence of probabilities such that the distributions $\mathbf{P}_{p_n}$ on Ω_n are asymptotically non-singular to the uniform distribution. Then $\mathcal{G}$ is uniformly noise-stable w.r.t. the distributions $\mathbf{P}_{p_n}$.*

The stability of uniform noise stability extends to small perturbations of the voters' profile distribution, and the distribution describing the noise does not require the assumption of independent voters.

We will now briefly deal with three or more alternatives. Let η be a probability distribution on order relations on a set A of alternatives. Given a set of n voters we will denote by $\mathbf{P}_\eta$ the probability distribution on the voters' profile where the voters' preferences are independent and are described by η. We will assume that for every two alternatives a and b the probability that aRb when R is chosen at random according to η is strictly between 0 and 1. We denote this probability by $\eta(a,b)$. We will study the distribution of social preferences when the voters' profiles are distributed according to $\mathbf{P}_\eta$. Consider the probability distribution p_η defined on asymmetric relations as follows:

$$p_\eta(R) = \prod_{aRb} \eta(a,b) \Big/ \sum_R \prod_{aRb} \eta(a,b). \tag{4.1}$$

This distribution $p_\eta(R)$ represents the following scenario. For every individual the order relations among the m alternatives are distributed according to η and are independent. The distribution on the social preference is based on the following procedure. For every two alternatives a and b choose an individual $v_{a,b}$ and let the the social preference between a and b agree with the preference of this individual. Assume that all these individuals are different. The resulting distribution on the social preferences is p_η.

We will now extend the definition of social chaos to strong monotone simple games and arbitrary distributions η on the individual order relations.

Definition 4.6. A sequence $(G_k)_{k=1,2,\ldots}$ of strong simple games leads to *social chaos* w.r.t. the probability distribution η on voters' order preferences if for every three alternatives a, b, and c the probability that the society prefers alternative a to alternative c given that it prefers a to b and b to c is asymptotically the same as the a priori probability that the society prefers a to c.

Given these definitions, Theorems 1.3, 1.8, 1.7 extend. We need to replace the uniform distribution on all asymmetric relations by the distribution p_{η_k}.

We conclude this section by a brief discussion of the assumption of small bias. The assumption of small bias is realistic and this by itself can be regarded as a surprising phenomenon. Why is it the case that so often shortly before an election we can give substantial probability for each one of two candidates to be elected? How come the probabilities that we can assign to the choices of each voter do not "sum up" to a decisive collective outcome? This seems especially surprising in view of the property of aggregation of information. It is worth noting that there are several factors that can push the situation of collective choice towards small bias:

(i) Abstention and other strategic considerations of voters: If a voter's incentive to vote is based (as is often assumed) on his chances of being pivotal, then for voters acting strategically, abstention or applying mixed strategies and not acting just based on the signals may drive the situation towards criticality. See, e.g., Feddersen and Pesendorfer (1996, 1998) where supermajority is studied, and Samet (2004) for general simple games.[7]

If there are two main candidates and a few additional candidates who are each ideologically close to one of these two, a voter may choose to vote for a minor candidate if he knows that his vote is not needed to decide between the main candidates. Again, this can push the situation between the two main candidates towards criticality.

(ii) Strategic considerations of candidates: In elections, such as those in the U.S., the candidates concentrate their efforts on swing states. This may lead to critical behavior in some of these states.

(iii) Choice out of equilibrium: Consider the difference between the following two scenarios: a) choosing between two types of products, b) choosing between two types of products when the difference in quality is reflected in their prices. In the second scenario it is more likely that we will observe critical behavior and even more so when the decision-maker's information about the difference in quality is based primarily on the difference in price.

(iv) Dependence: A major reason why the assumption of small bias is realistic is the lack of independence between voter preferences. Even a short

[7] In these cases strategic behavior leads to asymptotic complete aggregation of information in the sense that if there are independent individual signals weakly biased toward the superior alternative then with probability tending to one the superior alternative will be chosen.

time before an election there is a nonnegligible probability that an event will influence a large number of voters in the direction of one candidate.[8]

5. Dependence

Following is an example that demonstrates some of the issues that arise when the voting rule is simple majority and voter behavior is not independent. Suppose that the society is divided into a communities of b voters each. The number of voters is thus $n = ab$, which we assume is an odd number. Each voter i receives an independent signal s_i, where $s_i = 1$ with probability $1/2$ and $s_i = 0$ with probability $1/2$. The voters are aware of the signals of the other voters in their community and are influenced by them. Let $q > 0$ be a small real number. A voter changes his mind if he observes a decisive advantage for the other candidate in his community, i.e., if he observes an advantage where the probability of observing such an advantage or a larger one, when voter behavior is independent and uniform, is at most q.

The election's outcome as a function of the original signals $s_1, s_2, \ldots, s_n$ can be described by a strong simple game which we denote by $G[a; b; q]$. (The example will also "work" if a fraction $(1 - \alpha)$ of the voters vote according to their signals and only some small fraction $\alpha > 0$ of voters change their minds as before.[9]

Let us examine the situation for a sequence (G_n) of such examples where the parameters a and b both tend to infinity, $n = ab$, q tends to zero, and $\sqrt{1/q} = o(\sqrt{n})$. (For example, take $a = b = \sqrt{n}$, and $q = n^{-1/4}$.) In this case, G_n exhibit noise sensitivity for (independent) small amounts of noise in the original *signals*. The outcomes of elections as a function of the individual signals is thus completely chaotic. Indeed, the outcomes of the elections are determined (with very high probability) by the number of communities with decisive signals in favor of one of the candidates. The expected number of decisive communities is qa. The difference between the number of decisive communities favoring the two candidates behaves like $\sqrt{qa}$. Therefore, these

[8] This type of dependence can imply failure of aggregation of information in the following sense: it is possible that every voter's signal will favor a certain candidate with probability (say) 0.6 and yet there is a substantial probability that the other candidate will be elected regardless of the number of voters.

[9] Another variant that will have similar properties is the case in which there are election polls in each community that influence some small fraction of voters.

communities contribute $b\sqrt{qa}$ votes to one of the candidates. Since the original gap in the number of favorable signals between the two candidates behaves like $\sqrt{ab}$ and by our assumption $\sqrt{an} = o(n\sqrt{qa})$ we conclude that the decisive communities determine the outcome of the elections with very high probability. From this point on, the analysis is similar to the proof of Theorem 3.5.[10]

On the other hand, this same sequence is extremely noise-stable for independent noise with respect to counting the votes! The gap between votes cast for the two candidates behaves like $b\sqrt{qa}$ so that even if a random subset of 40% of the votes are miscounted the probability that the election's outcome will be reversed is extremely small.

The two properties of our example: The first property of chaotic behavior for noise affecting the original signal and the second property of strong stochastic stability for noise affecting individual votes seem characteristic to situations in which voter behavior depends on independent signals in a way that creates positive correlation between voters. This topic deserves further study. Note that when we consider random independent noise in the original signals, the distribution of the resulting votes is identical to the original distribution without the noise. This is not the case for random independent noise in counting the votes.

We now move from examples to more general models. A general model for voter behavior will be of a pair $\langle G, \nu \rangle$ where G is a strong simple game representing a voting rule and ν is a probability distribution of voters' preferences or signals. In this case we can consider random independent noise that can represent mistakes in counting the votes. For such a form of noise the distribution of the noisy signals will be different than the original distribution.

Another form of noise can be described as follows. Denote by $d(x, y)$ the Hammming distance between two vectors $x, y \in \Omega_n$. Given a vector $(x_1, \cdots, x_n)$ of voters' signals and another vector $y = (y_1, \ldots, y_n)$ define

$$q(y) = (1-t)^{n-d(x,y)} t^{d(x,y)} \cdot \nu(y),$$

and normalize $q(y)$ to a probability distribution by setting

$$p(y) = q(y) \Big/ \sum \{ q(z) \, : \, z \in \Omega_n \}.$$

[10] If $q = t/\sqrt{a}$, where t is small but bounded away from zero, then for every amount ε of noise in the signals, the correlation between the outcomes before and after the noise will tend to a small constant; in this case the situation depends on the original signals in a very, but not completely, chaotic manner.

If the noise changes x to y with probability $p(y)$ the distribution of the noisy voters' signals is the same as the original distribution. This type of noise can be regarded as applying to the (unknown) mechanism leading to the voters' preferences distribution ν.

A somewhat more restrictive but still quite general probabilistic setting for aggregation is the following: there are some independent signals $s_1, s_2, \ldots, s_t$ that may depend on the state of the world. At the stage where individuals make up their mind they are exposed to some of these signals directly and also to the emerging preferences of other individuals. The individual's final preferences are then aggregated according to some voting rule. When we study the sensitivity to noise it can make a big difference if we consider noise affecting the original signals or noise affecting the individual choices (mistakes in counting the votes).

In such a general framework there are cases, such as the example presented at the beginning of this section, when rather general simple games describe the situation even for a standard voting rule. For simplicity we can restrict ourselves to the case of simple majority rule. Suppose that the choice of the i-th individual $v_i = v_i(s_1, \ldots, s_t)$ is a function of the signals $s_1, \ldots, s_t$. If the v_i's are monotone and satisfy $v_i(1-s_1, 1-s_2, \ldots, 1-s_t) = 1-v_i(s_1, s_2, \ldots, s_t)$ then the outcome of the elections in terms of $s_1, s_2, \ldots, s_t$ is expressed by a proper simple game G (see also Kalai (2004), Section 3). If the signals are random and uniformly distributed on Ω_t then the situation can be considered within our frameworks. As we have shown, it is possible to provide natural examples in which the elections outcome as a function of the original signals will exhibit chaotic behavior. In addition to examples that are hierarchical, like those we have considered, one can also consider cases in which voters repeatedly update their positions based on the positions of friends and neighbors.

6. Abstention

In real elections it is common for a large proportion of voters not to vote. Abstention is thus the most common form of noise when it comes to real-life elections and perhaps also other forms of aggregation. It is possible to extend our model and consider monotone and neutral social welfare functions that allow individual indifference between alternatives. When there are two alternatives a and b, a social welfare function can

be described by a function $v(x_1, x_2, \ldots, x_n)$ where each $x_i \in \{-1, 0, 1\}$ and also $v(x_1, x_2, \ldots, x_n) \in \{-1, 0, 1\}$. Here $x_i = 1, 0, -1$ according to whether the i-th individual prefers a to b or is indifferent between the two, or prefers b to a, respectively. For two alternatives neutrality is equivalent to $v(-x_1, -x_2, \ldots, -x_n) = -v(x_1, x_2, \ldots, x_n)$. When we consider only ± 1 variables, the function v describes a proper simple game G.[11] We assume that v is monotone, namely, if $x_i \geq y_i, i = 1, 2, \ldots, n$ then $v(x_1, x_2, \ldots, x_n) \geq v(y_1, y_2, \ldots, y_n)$. The function v determines a neutral social welfare function on every set A of m alternatives.

We can study the stochastic behavior of social preferences as we did before. Let (v_n) be a sequence of functions of the types we considered above and let (G_k) be the proper simple game associated to (v_k). We need the assumption that for every $\alpha > 0$ if a random fraction of α among voters are indifferent between the two alternatives (in other words, if they abstain) and the voters that do not abstain vote randomly and uniformly, then the probability of social indifference tends to zero as n tends to infinity. This implies, in particular, that the sequence (G_k) has strongly diminishing bias.

Suppose that all voters vote; then we are back in the case we studied in this paper (with the assumption of strongly diminishing bias) and the notion of noise sensitivity and social chaos extend unchanged.

Our first theorem asserts that if (G_k) is a noise-sensitive sequence of proper monotone games then the effect of abstention of even a small fraction of voters is dramatic. (Note that the game G_k does not determine the function v_k but the result applies simultaneously for all neutral monotone extensions of the games G_k to cases of individual indifference.)

Consider the following scenario. Let R be the outcome of an election between two candidates and random voter profile $(x_1, x_2, \ldots, x_n)$ without abstention. Now consider another random voter profile $(y_1, y_2, \ldots, y_n)$ where $y_i = x_i$ with probability $1 - s$ and $y_i = 0$ with probability s. Let R' be the outcome of the election for the voter profile $(y_1, y_2, \ldots, y_n)$.

Theorem 6.1. *Consider a sequence (v_k) of monotone functions describing a neutral voting rule with individual indifference. Suppose that the social indifference is strongly diminishing and that the sequence (G_k) of associated proper games is noise-sensitive. Then the correlation between R and R' tends to zero as k tends to infinity.*

[11] Note that the payoff function v uses the values $-1, +1$ instead of $0, 1$ to describe this game.

Remark. As before, when we refer to "the correlation between R and R'" we mean "the correlation between the two events":

1. R represents the social preferences for $(x_1, \ldots, x_n)$
2. R' represents the social preferences for $(y_1, \ldots, y_n)$.

We abuse notation similarly below. Our next result asserts that the property of noise sensitivity is preserved with probability 1 under abstention.

Next consider the following scenario. Let R be the outcome of an election between two candidates and random voter profile $(x_1, x_2, \ldots, x_n)$ conditional on the assumption that the probability of a voter abstaining is $\alpha \geq 0$. Let $(y_1, y_2, \ldots, y_n)$ be another random voter profile where $y_i = x_i$ with probability $1 - s$ and $y_i = 0$ with probability s. Let R' be the outcome of the election for the voter profile $(y_1, y_2, \ldots, y_n)$.

Theorem 6.2. *Consider a sequence (v_k) of functions describing a monotone voting rule with individual indifference. Suppose that the social indifference is strongly diminishing and that the sequence (G_k) of associated proper games is noise-sensitive. Then the correlation between R and R' tends to zero as k tends to infinity.*

The proofs of these results are omitted.

7. Discussion

1. The superiority of majority. What are the simple games most stable under noise? It was conjectured by several authors that under several conditions which exclude individual voters having a large power, majority is (asymptotically) most stable to noise. (See a discussion and a surprising application to computer science in Khot, Kindler, Mossel and Ryan (2004)). This conjecture was recently proved by Mossel, O'Donnell and Oleszkiewicz (2005).

Recall that The sequence (G_k) of monotone simple games has *diminishing Banzhaf individual power* if

$$\lim_{n \to \infty} b_{\max}(G_k) = 0.$$

Theorem 7.1 (Mossel, O'Donnell and Oleszkiewicz (2005)). *For a sequence (G_n) of strong simple games with diminishing individual Banzhaf power,*

$$N_s(G_n) \geq \big(1 - o(1)\big) \cdot \left(\frac{\arccos(1-2s)}{\pi}\right).$$

In other words, when the individual Banzhaf power indices diminish simple majority is asymptotically most stable to noise.[12] A consequence of this theorem is

Corollary 7.2 (Mossel, O'Donnell and Oleszkiewicz (2005)). *For every sequence (G_n) of strong simple games with diminishing individual Banzhaf power,*

$$p_{\mathbf{cyc}}(G_n) \geq 1/4 - \big(3/(2\cdot\pi)\big) \cdot \arcsin(1/3) + o(1).$$

In other words, for social welfare functions with diminishing Banzhaf value, simple majority is asymptotically least likely to yield a cyclic social preference relation on three alternatives.

2. Aggregation of information. Given a strong monotone simple game G on a set V of players we denote by $\mathbf{P}_p(G)$ the probability that a random set S is a winning coalition when for every player $v \in V$ the probability that $v \in S$ is p, independently for all players. Condorcet's Jury theorem asserts that for the sequence G_n of majority games on n players

$$\lim_{n\to\infty} \mathbf{P}_p(G_n) = 1, \quad \text{for every} \quad p > 1/2. \tag{7.1}$$

This result, a direct consequence of the law of large numbers, is referred to as asymptotically complete aggregation of information. In Kalai (2004) it was proved that for a sequence of monotone simple games asymptotically complete aggregation of information is equivalent to diminishing Shapley–Shubik individual power.

The speed of aggregation of information for simple majority games is described by the central limit theorem. For a monotone strong simple game G, let $T_\varepsilon(G) = [p_1, p_2]$ and $t_\varepsilon(G) = p_2 - p_1$, where $\mathbf{P}_{p_1}(G) = \varepsilon$ and $\mathbf{P}_{p_2}(G) = 1 - \varepsilon$. When G_n is simple majority on n players the central

[12] It is crucial to fix s and let n tend to infinity. If s tends to zero as n tends to infinity the situation can be very different. For example, when $s = 1/n$ the majority function is most sensitive to noise.

limit theorem implies that $t_\varepsilon(G_n) = \theta\big(1/\sqrt{n}\big)$. It is known that for n-player games (n being an odd integer) and every ε, $t_\varepsilon(G)$ is minimized by the simple majority game[13].

The next theorem asserts that uniform noise stability (in a slightly stronger sense than before) implies that up to a multiplicative constant the aggregation of information is optimal.

Theorem 7.3. *Let (G_k) be a sequence of strong simple games such that G_k has n_k players and let $\varepsilon > 0$ be a fixed real number. Suppose that the class of games and distributions $(G_k, \mathbf{P}_p)$, $p \in T_\varepsilon(G_k)$ is uniformly noise-stable. Then $t_\varepsilon(G_n) = \theta\big(1/\sqrt{n_k}\big)$.*

3. Indeterminacy. Consider a sequence (G_n) of strong simple games. Let $m \geq 3$ be a fixed integer, let A be a set of m alternatives, and denote by F_n the neutral social welfare function based on G_n when the set of alternatives is A. Let R be an arbitrary asymmetric relation on A. It was proved in Kalai (2004) that if (G_n) has diminishing individual Shapley–Shubik power then, for n sufficiently large, R is in the range of F_n (in other words, $p_R(G_n) > 0$) and this property is referred to there as *complete indeterminacy*. Note that social chaos (in view of Theorem 1.7) can be regarded as the ultimate form of social indeterminacy as it implies that all the probabilities $p_R(G_n)$ are asymptotically the same.

The main tool for showing that diminishing individual Shapley–Shubik power implies social indeterminacy is the equivalence between diminishing individual Shapley–Shubik power and asymptotically complete aggregation of information. It also follows from the results of Kalai (2004) that diminishing individual Shapley–Shubik power implies diminishing individual Banzhaf power (but not vice versa). When we assume diminishing individual Banzhaf power we can expect the following strong form of complete indeterminacy.

Definition 7.4. A sequence (G_n) of strong simple games leads to *stochastically complete social indeterminacy* if for every asymmetric relation R on a set A of m alternatives

$$\liminf_{n\to\infty} p_R(G_n) > 0.$$

[13] The fact that the derivative $d\mathbf{P}_p(G)/dp$ is maximized for simple majority can be found, e.g., in Friedgut and Kalai (1996).

Theorem 7.5. (1) *A sequence (G_n) of strong simple games with diminishing Banzhaf individual power leads to stochastically complete social indeterminacy.*

(2) *A sequence (G_n) of strong simple games that is uniformly chaotic leads to stochastically complete social indeterminacy.*

The proof of Theorem 7.5 uses the argument used in proving Theorem 1.7 and for part (1) we require also a recent theorem of Mossel, O'Donnell and Oleszkiewicz (2005), referred to as "it ain't over until it's over". This theorem roughly asserts that for any random voter profile when a large random set of votes is counted there is still (almost surely as n tends to infinity) a probability bounded away from zero (however tiny) that either candidate will win. We defer further details to the Appendix.

4. Noise stability. This paper is mainly devoted to the chaotic behavior of noise-sensitive classes of simple games. There are interesting issues concerning the behavior of social preferences for social welfare functions based on uniformly noise-stable classes of simple games, and random uniform voter profiles. I will mention one topic worth further study: suppose that we are interested in the social preference relation between two alternatives A and B but we can cannot compare these two alternatives directly. Rather, we can compare both A and B with other alternatives $C_1, C_2, \dots C_r$. Suppose that for all those alternatives we discover that the society prefers A to C_i and C_i to B. It seems true that as r grows we can conclude that the society prefers A to B with higher and higher probability that tends to 1 as r tends to infinity. This is unknown even for simple majority. (In contrast, it follows from Theorem 1.7 that for every fixed r and a noise-sensitive sequence (G_n) of strong simple games, the a posteriori probability that the society prefers A to B tends to $1/2$.)

5. Other economic models. We study noise sensitivity and its chaotic consequences for social welfare functions. It would be interesting to study similar questions for other economic models and, in particular, for exchange economies and more general forms of economies. Related notions of indeterminacy, aggregation of information, pivotal agents, and power were considered for various models of economies and some analogies with results in social choice theory can be drawn. The well-known Sonnenschein–Debreu–Mantel theorem concerning demand functions for exchange economies is an indeterminacy result that implies that various dynamics for reaching equilibrium points can be chaotic. Yet, it appears that exchange economies

behave in a similar way to weighted majority functions and are thus quite stable to noise in terms of initial endowments and individual demand functions. More complicated models of economies (or simple models under a complicated probabilistic environment) appear to exhibit a more chaotic behavior in the sense of this paper.

6. The asymptotic nature of our results. We would like to stress that our approach and results are asymptotic: we consider the situation when the number of individuals tends to infinity. The number of alternatives m and the noise ε are supposed to be fixed as n tends to infinity, so one has to be careful in drawing conclusions for numerical values of n, ε, and m. Moreover, the dichotomy between stochastic stability and complete chaos is based on first fixing ε and letting n tend to infinity and then letting ε tend to zero. Studying the dependence on ε even for multi-level majority-based rules gives a more involved picture that we briefly discussed in Section 3.2.

8. Conclusion

Social welfare functions form a simple and basic model in economic theory and political science. We have considered their probabilistic behavior under uniformly distributed voter profiles. In contrast to social welfare functions based on simple majority (or weighted majority) which demonstrate stable behavior to small stochastic perturbations, in this paper we studied a class of social welfare functions that demonstrate completely chaotic behavior and showed that this class can be characterized in surprisingly different ways. This class contains some rather natural examples. Chaotic behavior occurs in our model for hierarchical or recursive aggregation of preferences and in cases where the social preferences heavily depend on a small number of small communities with decisive views.

One interpretation of our results relates to the primary role of the majority rule. There have been many efforts to demonstrate the dominance of majority among voting rules. Analysis of sensitivity to noise gives a clear advantage to the majority rule. Our results on the stochastic robustness of majority and the chaotic nature of methods that are asymptotically orthogonal to majority point in this direction. So is the result by Mossel, O'Donnell and Oleszkiewicz (2005) on the asymptotic optimality of the majority rule in terms of sensitivity to noise. The majority rule is a very simple and very basic economic/political mechanism and an analysis of

sensitivity to noise in a similar manner to ours may be an important aspect in evaluating other economic mechanisms.

Do we witness chaotic behavior in realistic economic situations of preference aggregation? A complete social chaos of the kind considered in this paper is an ideal rather than a realistic economic phenomenon. However, I would expect to find chaotic components in real-life examples of preference aggregation and in my opinion complete (stochastic) stability of the kind we observe for weighted majority is often unrealistic. Understanding the chaotic components (in the sense of this paper) in other stochastic economic models is worthy of further study.

9. Appendix

9.1. Proofs of properties of simple majority

Proof of Theorem 1.6. Let f_{2k+1} be the majority function on the set $V = \{1, 2, \ldots, 2k+1\}$. In order to prove Lemma 3.1 we will study the Fourier–Walsh coefficients of f_{2k+1}. For a Boolean function f the coefficient $\widehat{f}(\{j\})$ by definition equals $\sum_{S \subset V,\ j \notin S} 2^{-n} f(S) - \sum_{S \subset V,\ j \in S} 2^{-n} f(S)$. Therefore, if f is monotone we obtain that

$$\widehat{f}(\{j\}) = -b_j(f)/2,$$

where $b_j(f)$ is the Banzhaf value of j in f.

For f_{2k+1} we obtain that $\widehat{f}(\{j\}) = \binom{2k}{k}/2^{2k+1}$. Recall that for a Boolean function f and for $k \geq 0$ we denoted

$$W_k(f) = \sum \{ \widehat{f}^2(S) : S \subset N,\ |S| = k \}.$$

For every Boolean function that represents a strong simple game:

(1) $W_0(f) = \widehat{f}^2(\emptyset) = 1/4$,

(2) $\sum_{k=0}^{n} W_k^2(f) = \|f\|_2^2 = 1/2$, and

(3) $W_k(f) = 0$ for an even integer $k > 0$ (Proposition 2.1).

For simple majority on $2k+1$ voters we obtain that

$$W_1(f) = \sum_{j=1}^{2k+1} \hat{f}^2(\{j\}) = (2k+1)\binom{2k}{k}/2^{2k+1}.$$

It is easy to verify that this expression is monotone decreasing with k. Therefore, if $k \geq 1$ the value of $W_1(f_{2k+1})$ is maximized for $k = 1$, namely, for simple majority on three voters. For $k = 1$ we get $W_1(f_3) = 3/16$. Of course, $\hat{f}_3(S) = 0$ for $|S| > 3$ and therefore $W_k(f_3) = 0$ for every $k > 3$. It follows that for every $k \geq 1$, $W_1(f_{2k+1}) \leq W_1(f_3)$ and $W_r(f_{2k+1}) \geq W_r(f_3)$.

These relations with Theorems 2.3 and 2.2 imply our result. Indeed, Theorem 2.3 asserts that the probability $N_\varepsilon(f)$ that an ε-noise will change the outcome of an election is given by the formula:

$$N_\varepsilon(f) = 1/2 - 4 \cdot \sum_{r=1}^{n} (1-\varepsilon)^r W_r(f).$$

When we compare this expression for f_3 and f_{2k+1} the result follows from the fact that when we pass from f_3 to f_{2k+1} the weights W_i's (whose sum is constant) are shifted from the lower values of i to the larger ones. Formally, the proof is completed by the following standard lemma:

Lemma 9.1. *Let $a_1 > a_2 > \ldots, a_n > 0$ be a sequence of real numbers. Let $b = (b_1, b_2, \ldots, b_n)$ and $b' = (b'_1, b'_2, \ldots, b'_n)$ be sequences of positive real functions such that $\sum_{i=1}^n b_i = \sum_{i=1}^n b'_i$ and for every r, $1 \leq r \leq n$, $\sum_{i=1}^r b_i \geq \sum_{i=1}^r b'_i$ then $\sum_{i=1}^n a_i b_i \geq \sum_{i=1}^n a_i b'_i$.*

The same argument or just applying Relation 2.7 implies part (i).

9.2. Proof of Theorem 4.3

Theorem 4.3 is the only result in this paper that requires a non-elementary tool from harmonic analysis. We will give the proof for the case where $p_k = 1/2$, namely, for the uniform probability distribution. In this case the bias comes from the voting rule and not from the probability distribution. Corollary 2.4 (which is quite elementary) extends to the case of Boolean functions that may have large bias, as follows:

Theorem 9.2. *A sequence (f_n) of Boolean functions is noise-sensitive if and only if, for every $k > 0$,*

$$\lim_{n\to\infty} \sum_{i=1}^{k} W_i(f_n) / \left(\sum_{i\geq 1} W_i(f_n) \right) = 0. \tag{9.1}$$

Note that by the Parseval formula $\sum_{i\geq 1} W_i(f_n) = q_n(1-q_n)$, where $q_n = \|f_n\|_2^2$ is simply the probability that f_n equals 1. (In the case that (f_n) represent a strong simple game or represent games with small bias the denominator on the right-hand side of relation (9.1) is bounded from zero and can be deleted.) In view of formula (9.1) relating the noise sensitivity to the Fourier coefficients, in order to demonstrate noise sensitivity we have to show that most of the Fourier coefficients of f_n (in terms of the 2-norm of f_n) are concentrated on large "frequencies".

For a real function $f : \Omega_n \to \mathbb{R}$, $f = \sum \widehat{f}(S) U_S$, define

$$T_\varepsilon(f) = \sum \widehat{f}(S) \varepsilon^{|S|} U_S.$$

The Bonamie–Beckner inequality (see, e.g., Benjamini, Kalai and Schramm (1999)) asserts that for every real function f on $\Omega(n)$,

$$\left\| T_\varepsilon(f) \right\|_2 \leq \|f\|_{1+\varepsilon^2}.$$

For a Boolean function f, note that for every t, $\|f\|_t^t = \mathbf{P}(f)$. It follows from the Bonamie–Beckner inequality (say, for $\varepsilon = 1/2$) that for some constant K,

$$\sum_{0<|S|<\log(1/t)/4} \widehat{f}^2(S) \leq K \sqrt{\left(q_n(1-q_n)\right)}.$$

This is sufficient to show that relation 9.1 holds even when k depends on n and is set to be $(1/4) \cdot \log\left(1/\left(q_n \cdot (1-q_n)\right)\right)$.

9.3. It ain't over until it's over

Sketch of the proof of Theorem 7.5. In the proof of Theorem 1.7 we considered two asymmetric relations R and R' on m alternatives that differ only in the ranking of the two last alternatives and used noise sensitivity to show that the ratio $p_R(G_n)/p_{R'}(G_n)$ tends to one as n tends to infinity.

If we want to show only that the ratio $p_R(G_n)/p_{R'}(G_n)$ is bounded away from 0 we need a weaker property for the sequence (G_n), referred to as [IAOUIO], which is defined as follows:

Let G be a monotone simple strong game considered as a voting rule between Alice and Bob with n voters. Let $\varepsilon > 0$ and $\delta > 0$ be small real numbers. Given a set S of voters and the way these voters vote, let Q denote the probability when the remaining voters vote uniformly at random that Alice will win the elections. Let $q(G; \varepsilon, \delta)$ be the probability that a random choice of a set S of $(1-\varepsilon)n$ of the voters and a random choice of the way voters in S voted the value of Q is at least δ.

Consider a sequence (G_n) of strong simple games regarded as voting rules. The sequence (G_n) has property [IAOUIO] if

- For every $\varepsilon > 0$ there is $\delta > 0$ so that

$$\lim_{n\to\infty} q(G; \varepsilon, \delta) = 1.$$

It can be shown using the central limit theorem that the sequence of simple majority games on n players has the property [IAOUIO]. In this case δ can be taken to be $e^{-\varepsilon^{-2}}$. As indicated above the proof of Theorem 1.7 gives

Theorem 9.3. *A sequence (G_n) of strong simple games with property* [IAOUIO] *leads to stochastically complete social indeterminacy.*

Showing property [IAOUIO] for a uniformly chaotic sequence of games is immediate from the definition and therefore part (2) of Theorem 7.5 follows directly from the proof of Theorem 1.7. To prove the first part we need to show that diminishing Banzhaf individual power implies the property [IAOUIO]. This is much more difficult and was recently proved by Mossel, O'Donnell and Oleszkiewicz (2005) in response to a conjecture opsed by the author and Friedgut.

Theorem 9.4 (It ain't over until it's over [Mossel, O'Donnell and Oleszkiewicz]). *A sequence of strong monotone simple games (G_n) with diminishing individual Banzhaf power has property* [IAOUIO].

REFERENCES

[1] K. Arrow, A difficulty in the theory of social welfare, *J. of Political Economy,* **58** (1950), 328–346.

[2] C. E. Bell, A random voting graph almost surely has a Hamiltonian cycle when the number of alternatives is large, *Econometrica,* **49** (1981), 1597–1603.

[3] I. Benjamini, G. Kalai and O. Schramm, Noise sensitivity of Boolean functions and applications to percolation, Publ. I.H.E.S. (1999).

[4] T. Feddersen, and W. Pesendorfer, The swing voter's curse, *The American Economic Review,* **86** (1996), 408–424.

[5] T. Feddersen, and W. Pesendorfer, Convicting the innocent: the inferiority of unanimous jury verdicts under strategic voting, *The American Political Science Review,* **92** (1998), 23–35.

[6] P. C. Fishburn, W. V. Gehrlein and E. Maskin, Condorcet proportions and Kelly's conjectures, *Discrete Appl. Math.,* **1** (1979), 229–252.

[7] T. E. Friedgut, *Political Participation in the USSR,* Princeton University Press, Princeton, N.J. (1979).

[8] E. Friedgut, and G. Kalai, Every monotone graph property has a sharp threshold, *Proc. American Mathematical Society,* **124** (1996), 2993–3002.

[9] W. V. Gehrlein, Condorcet's paradox and the Condorcet efficiency of voting rules, *Math. Japon.,* **45** (1997), 173–199.

[10] R. Holzman, E. Lehrer and N. Linial, Some bounds for the Banzhaf index and other semivalues, *Math. Oper. Res.,* **13** (1988), 358–363.

[11] Kalai, G., A Fourier-theoretic perspective for the Condorcet paradox and Arrow's theorem, *Adv. in Appl. Math.,* **29** (2002), 412–426

[12] G. Kalai, Social indeterminacy, *Econometrica,* **72** (2004), 1565–1581.

[13] S. Khot, G. Kindler, E. Mossel and R. O'Donnell, Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? *SIAM Journal of Computing,* **37(1)** (2007), pp. 319–357. (Also FOCS 2004).

[14] E. Mossel and R. O'Donnell, On the noise sensitivity of monotone functions, *Random Structures Algorithms,* **23** (2003), 333–350.

[15] E. Mossel, R. O'Donnell and K. Oleszkiewicz, Noise stability of functions with low influence: invariance and optimality, *Annals of Mathematics,* to appear 2010. (Also FOCS 2005.)

[16] Y. Peres, *Noise stability of weighted majority,* preprint math.PR/0412377 (2004).

[17] D. G. Saari, *Chaotic elections! A mathematician looks at voting,* American Mathematical Society, Providence, R I. (2001).

[18] Y. Samet, *Equilibria with information aggregation in sharp threshold voting rules,* M.Sc. thesis, Jerusalem, 2004.

[19] W. Sheppard, On the application of the theory of error to cases of normal distribution and normal correlation, *Phil. Trans. Royal Soc. London,* **192** (1899), 101–168.

Gil Kalai
Hebrew University of Jerusalem
and
Yale University
e-mail: kalai@math.huji.ac.il

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 213–238.

Coloring Uniform Hypergraphs with Small Edge Degrees

ALEXANDR V. KOSTOCHKA*, MOHIT KUMBHAT and VOJTĚCH RÖDL†

Dedicated to the 60th birthday of László Lovász

Let k be a positive integer and $n = \lfloor \log_2 k \rfloor$. We prove that there is an $\varepsilon = \varepsilon(k) > 0$ such that for sufficiently large r, every r-uniform hypergraph with maximum edge degree at most

$$\varepsilon(k)\, k^r \left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}}$$

is k-colorable.

1. Introduction

The degree of an edge e in a hypergraph G is the number of other edges of G intersecting e, and *the maximum edge degree* of G is the maximum over the degrees of its edges. A natural question is: Which bound on the maximum edge degree of an r-uniform hypergraph G provides that G is k-colorable? The classical result in this direction belongs to Erdős and Lovász. In their seminal paper [2] (where the Lovász Local Lemma appeared), they proved the following bound.

*Research of this author is supported in part by NSF grant DMS-0650784 and by grant 06-01-00694 of the Russian Foundation for Basic Research.

†Research of this author was supported by NSF grants DMS-0300529 and DMS-0800070.

Theorem 1 [2]. *If $k, r \geq 2$, then every r-uniform hypergraph with maximum edge degree at most $\frac{1}{4}k^r$ is k-colorable.*

The proof works also for list coloring. A remarkable feature of this result is that it works for all $k, r \geq 2$, and in many cases the bound is rather close to the best possible. In particular, Erdős and Lovász [2] showed that the bound cannot be significantly improved even if we consider only hypergraphs with high girth. If we denote by $D(k,r)$ (respectively, $D(k,r,g)$) the minimum D such that there exists an r-uniform non-k-colorable hypergraph G with maximum degree D (and girth at least g), then the results mentioned above can be summarized as follows. For every $g \geq 2$,

$$\frac{1}{4}k^r < D(k,r) \leq 20r^3k^{r-1}. \tag{1}$$

The upper bound in (1) was recently improved for $k < r$ by Kostochka and Rödl [5] to $r\lceil r\, k^{r-1} \ln k\rceil$.

Let $m(r,k)$ denote the minimum number of edges in an r-uniform hypergraph that is not k-colorable. Elaborating the proof of a lower bound on $m(r,2)$, and using the Local Lemma, Radhakrishnan and Srinivasan [6] improved the lower bound on $D(2,r)$ for large r to

$$D(2,r) \geq 0.17 \cdot 2^r \sqrt{r/\ln r}. \tag{2}$$

The main result of this paper is the following extension of (2) to fixed k and large r.

Theorem 2. *For every integer $k \geq 2$, let $\varepsilon = \varepsilon(k) = \exp\{-4k^2\}$ and $n = n(k) = \lfloor \log_2 k \rfloor$. Then for every sufficiently large r, every r-uniform hypergraph with maximum edge degree at most $D = \varepsilon k^r \left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}}$ is k-colorable. In other words, $D(k,r) > \varepsilon k^r \left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}}$.*

Recall that for the class of simple hypergraphs, Kostochka and Kumbhat [4] recently proved a slightly stronger bound than Theorem 2 gives. Recall that a hypergraph G is *b-simple* if $|e \cap e'| \leq b$ for every distinct $e, e' \in E(G)$.

Theorem 3 [4]. *If $b \geq 1$, $k \geq 2$, and $\varepsilon > 0$ are fixed and r is sufficiently large, then every r-uniform b-simple hypergraph G with maximum edge degree at most $k^r r^{1-\varepsilon}$ is k-colorable.*

We think that the statement of Theorem 3 holds for all r-uniform hypergraphs.

The proof of Theorem 2 uses the Local Lemma, ideas of Radhakrishnan and Srinivasan [6], and the proof of the lower bound on $m(r,k)$ by Kostochka [3].

The structure of the paper is as follows. In the next section, a semi-random procedure Evolution is described and some of its simple properties are derived. In Section 3 we study the structure of so called *cause trees* arising in the analysis of Evolution. In the next two sections we define some auxiliary "bad" events and estimate their probabilities. Using the independence structure of these auxiliary events and the Local Lemma, in the final section we show that for hypergraphs satisfying the conditions of Theorem 2, with positive probability Evolution gives a proper k-coloring. This means that such a coloring exists.

2. Coloring Procedure Evolution and its Properties

Let k, n and ε be as in the statement of the theorem. Let

$$r \geq \exp\left\{2\varepsilon^{-2(n+1)}\right\}. \tag{3}$$

Throughout the paper we will use the notation

$$z = \lfloor 4k^2 r/\ln r \rfloor. \tag{4}$$

Fix some $0 < p < 2^{-k^r r}$. Then there is the unique positive integer s such that $sp \leq \frac{\ln r}{(n+1)r} < (s+1)p$. Let $G = (V,E)$ be an r-uniform hypergraph with maximum edge degree at most $D = \varepsilon k^r \left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}}$.

The coloring procedure Evolution described below consists of $n+1$ stages, and every stage apart from Stage 0 consists of s steps. For $1 \leq l \leq n$ and $1 \leq i \leq s$, Step $(l-1)s+i$ is the ith step in Stage l.

We also fix a linear order L on $V(G)$. Now, the procedure works as follows.

Stage 0. (O1) Color every vertex $v \in V(G)$ randomly and independently, with a color $\phi(v) \in \{0,1,2,\ldots,k-1\}$ chosen uniformly in this set.

(O2) For every $v \in V(G)$, define the random variable $I(v)$ with the values in $\{1, 2, \ldots, sn\} \cup \{\infty\}$ as follows:

$$\mathbf{Pr}\{I(v) = x\} = \begin{cases} p, & \text{if } x \in \{1, 2, \ldots, sn\}; \\ 1 - psn, & \text{if } x = \infty. \end{cases} \tag{5}$$

Each random variable $I(v)$ is defined to be mutually independent of all other $I(w)$.

Stage l, $l = 1, \ldots, n$.
STEP $i + s(l-1)$, $1 \leq i \leq s$. Following order L, for one by one vertex $v \in V(G)$, check whether

(C1) $I(v) = (l-1)s + i$ and

(C2) v belongs to an edge that was monochromatic, say, of color α, before Stage l, and still is monochromatic at the current moment.

If both conditions (C1) and (C2) hold, then recolor v with color $\alpha + 2^{l-1}$ (modulo k). Otherwise, do nothing with v.

Remark 1. By Condition (C1), each vertex can be recolored at most once.

Remark 2. As it follows from the description of the procedure, every step consists of $\big|V(G)\big|$ smaller steps (one per vertex).

Lemma 1. *For every $w, q \geq 1$, every set $W \subseteq V$ with $|W| = w$, and every set $Q \subseteq \{1, 2, \ldots, sn\}$ with $|Q| = q$, the probability that for each vertex $v \in W$, $I(v) \in Q$ is at most $(qp)^w$.*

Proof. For every vertex $v \in V(G)$ and every $1 \leq l \leq n$ and $1 \leq i \leq s$, $\mathbf{Pr}\{I(v) = s(l-1) + i\} = p$. Therefore, the probability that $I(v) \in Q$ is at most qp. The mutual independence of all $I(v)$ yields the lemma. ■

For an edge $e \in E$ and $1 \leq l \leq n$, let

$$M(e, l) = \{v \in e : I(v) \leq sl\}. \tag{6}$$

Lemma 2. *For every $e \in E$ and $1 \leq l \leq n$,*

$$\mathbf{Pr}\{\big|M(e, l)\big| \geq z\} \leq \varepsilon^{0.5r}.$$

Proof. It is enough to prove the lemma for $l = n$. Observe that $|M(e,n)| \geq z$ means that there exists a set $Z \subset e$ with $|Z| = z$ such that $I(v) \leq sn$ for every $v \in Z$. Hence by Lemma 1 applied with $Q = [sn]$ and $W = Z$, this probability is at most

$$\binom{r}{z}(nsp)^z \leq \left(\frac{er}{z}\right)^z \left(\frac{n \ln r}{(n+1)r}\right)^z \leq \left(\frac{ne \ln r}{z(n+1)}\right)^z.$$

Since r is large and $z = \lfloor 4k^2 r / \ln r \rfloor > n$,

$$\frac{ne \ln r}{z(n+1)} \leq \frac{e \ln r}{z+1} \leq \frac{e \ln^2 r}{4k^2 r} \leq r^{-0.6}.$$

Thus

$$\left(\frac{ne \ln r}{z(n+1)}\right)^z \leq \left(r^{-0.6}\right)^{(4k^2 r/\ln r)-1} < e^{-2k^2 r} = \varepsilon^{0.5r}. \qquad \blacksquare$$

Lemma 3. *If a vertex is of color α at the end of Stage l, $l \geq 1$, then at the end of Stage 0 it can be colored only with colors $\alpha, \alpha - 2^0, \alpha - 2^1, \ldots, \alpha - 2^{l-1}$ (modulo k).*

Proof. By Remark 1, every vertex can be recolored at most once and by definition, a vertex of color β can be recolored during Stage j only with color $\beta + 2^{j-1}$ (modulo k). $\blacksquare$

Definition [Blaming edges]. If an edge e_0 becomes monochromatic of color α during Stage l, then it must contain at the end of Stage 0 a vertex of color $\alpha - 2^{l-1}$. Suppose that at the end of Stage 0 it contained vertices of colors $\alpha - 2^{l_1-1}, \ldots, \alpha - 2^{l_h-1}$, where $l_h = l$ and $l_1 < l_2 < \cdots < l_h$. Then for every $1 \leq j \leq h$, there exists an edge e_j and a vertex $v_j \in e_0 \cap e_j$ such that

(a) e_j was monochromatic of color $\alpha - 2^{l_j-1}$ at the end of Stage $l_j - 1$;

(b) v_j was recolored with α during Stage l_j and it was the last vertex of this color in e_0 recolored with α.

In this case we say that e_0 and v_j *l_j-blame* e_j.

Remark 3. If an edge e_0 becomes monochromatic of color α during Stage l, then all vertices of color $\alpha - 2^{l-1}$ in e_0 change their colors to α during this stage. But since in every step, the vertices of G are considered consecutively, there is the last vertex v^* of color $\alpha - 2^{l-1}$ in e_0 that changes its color. At that moment, any monochromatic edge e of color $\alpha - 2^{l-1}$ containing v^* shares only v^* with e_0.

Remark 4. It may happen that an edge e_0 can blame more than one edge containing the same vertex v_j. On the other hand, by definition, e_0 cannot blame an edge containing another vertex $v \in e_0$ with $\phi(v) = \phi(v_j)$.

Definition [Cause trees]. If an edge e_0 is monochromatic of color α at the end of Stage l, then a *cause tree* $T = T(e_0, \alpha, l)$ is a subset of edges of G defined by induction on l as follows:

(a) The set T always contains e_0.

(b) If e_0 was monochromatic of color α already after Stage 0, then $T = \{e_0\}$ for every l.

(c) Suppose that at the end of Stage 0 edge e_0 contained vertices of colors $\alpha - 2^{l_1-1}, \ldots, \alpha - 2^{l_h-1}$, where $l_h \le l$ and $l_1 < l_2 < \cdots < l_h$. Suppose further that for $j = 1, \ldots, h$, edge e_0 l_j-blames edge e_j. Then

$$T = T(e_0, \alpha, l) = \{e_0\} \cup \bigcup_{j=1}^{h} T(e_j, \alpha - 2^{l_j-1}, l_j - 1).$$

Remark 5. By Remark 4 and the definition of cause trees, it could be that in the same outcome of Evolution for the same triple (e_0, α, l), we can construct several distinct cause trees $T = T(e_0, \alpha, l)$.

Definition [Levels of edges]. If $T = T(e_0, \alpha, l)$ is defined as above, then we also say that $e_1, e_2, \ldots, e_h$ are the *edges of level* 1 of T, the edges blamed by the edges of level 1 are the *edges of level* 2 of T, and so on. Thus, if an edge e of a cause tree has vertices of exactly t distinct colors at the end of Stage 0, then e blames either $t-1$ or t other edges.

3. Structure of Cause Trees

Since each vertex can be recolored at most once, each edge at different stages of Evolution can become monochromatic with at most two colors. Furthermore, if an edge e was monochromatic of a color α_1 after Stage l_1 and becomes monochromatic of a color $\alpha_2 \neq \alpha_1$ after Stage l_2, then e has to be monochromatic of color α_1 already after Stage 0 and all vertices of e change their color to $\alpha_2 = \alpha_1 + 2^{l_2-1}$ at Stage l_2. In this case, each cause tree for e considered after Stage l_2 has exactly one edge of level 1.

In view of this, if an edge e becomes monochromatic exactly once during Evolution, then the corresponding color α is called *the main color of* e and denoted by $\mu(e)$, and if e becomes monochromatic twice, then *the main color of* e, $\mu(e)$, is the first of these two colors.

If e is monochromatic of some color α after some Stage l, then we say that e is an *l-unlucky edge.*

Lemma 4. *If e_0 is an l-unlucky edge with a cause tree T, then the main colors of all the edges of T are distinct.*

Proof. If e and e' are edges of T, then there exist two sequences $e_0, e_1, \dots, e_q = e$ and $e'_0 = e_0, e'_1, \dots, e'_{q'} = e'$ such that e_j l_j-blames e_{j+1} for $j = 0, 1, \dots, q-1$ and e'_j l'_j-blames e'_{j+1} for $j = 0, 1, \dots, q'-1$. Furthermore, $l_0 > l_1 > \dots > l_{q-1}$, $l'_0 > l'_1 > \dots > l'_{q-1}$, and the sequences $l_0, l_1, \dots, l_q$ and $l'_0, l'_1, \dots, l'_q$ are not identical. Thus, the numbers $2^{l_0-1}+2^{l_1-1}+\dots+2^{l_{q-1}-1}$ and $2^{l'_0-1}+2^{l'_1-1}+\dots+2^{l'_{q'-1}-1}$ are distinct and differ by less than k. On the other hand, by definition, the main color of e is $\alpha-2^{l_0-1}-2^{l_1-1}-\dots-2^{l_{q-1}-1}$ and the main color of e' is $\alpha - 2^{l'_0-1} - 2^{l'_1-1} - \dots - 2^{l'_{q'-1}-1}$. This proves the lemma. ■

Lemma 5. *Suppose that e_0 is an l-unlucky edge with a cause tree T. If e and e' are edges of T and neither of them blames the other, then e and e' are disjoint.*

Proof. Assume that e and e' have a common vertex v and both belong to T. Then there exist two sequences $e_0, e_1, \dots, e_q = e$ and $e'_0 = e_0, e'_1, \dots, e'_{q'} = e'$ such that e_j l_j-blames e_{j+1} for $j = 0, 1, \dots, q-1$ and e'_j l'_j-blames e'_{j+1} for $j = 0, 1, \dots, q'-1$. Furthermore, $l_0 > l_1 > \dots > l_{q-1}$, $l'_0 > l'_1 > \dots > l'_{q-1}$.

Claim 1. $l_{q-1} \neq l'_{q'-1}$.

Proof of Claim. If $l_{q-1} = l'_{q'-1}$, then e and e' both were monochromatic at the end of Stage $l_{q-1}-1$. But by Lemma 4, their main colors differ. This proves the claim.

Thus below we can assume that $l_{q-1} < l'_{q'-1}$. It follows that e ceased to be monochromatic before e' did. In particular, the vertex $v \in e \cap e'$ was recolored from $\mu(e)$ to $\mu(e')$. This yields that

(7) $\qquad \mu(e') - \mu(e)$ (modulo k) is a power of 2.

Claim 2. $\mu(e') - \mu(e) = 2^{l_{q-1}-1}$ modulo k.

Proof of Claim. Recall that

$$\mu(e') - \mu(e) = \left(\alpha - 2^{l'_0-1} - 2^{l'_1-1} - \cdots - 2^{l'_{q'-1}-1}\right) - \left(\alpha - 2^{l_0-1} - 2^{l_1-1} - \cdots - 2^{l_{q-1}-1}\right).$$

In this expression, α cancels out and every other summand apart from $2^{l_{q-1}-1}$ is divisible by $2^{l_{q-1}}$. Together with (7), this yields the claim.

Claim 2 implies that v was recolored during Stage l_{q-1} and thus $\mu(e') = \mu(e_{q-1})$. This contradicts Lemma 4. ■

Lemma 6. *Let e_0 be an edge of G that is unlucky after Stage l. Let T be a cause tree for this event such that for every pair of edges (e_1, e_2) of T such that e_1 blames e_2, we have $\mu(e_1) - \mu(e_2) \in \{1, 2, \ldots, 2^{l-1}\}$ (modulo k). Then $|E(T)| \leq 2^l$. In particular, each cause tree has at most $2^n \leq k$ edges.*

Proof. If e_1 l_1-blames e_2 and e_2 l_2-blames e_3, then $l_2 < l_1$. Thus, under conditions of the lemma, for the root e_0 and an arbitrary edge e of the tree, we have

$$\mu(e_0) - \mu(e) \in \left\{1, 2, \ldots, 2^{l-1} + 2^{l-2} + \cdots + 1\right\} = \{1, 2, \ldots, 2^l - 1\}.$$

Now, Lemma 4 implies that T has at most $1 + (2^l - 1)$ edges. ■

Below we will analyze which subsets of edges of G can form cause trees $T(e, \alpha, l)$ for some values of e, α and l. Lemma 5 implies that every cause tree $T = T(e, \alpha, l)$ is an r-uniform hypergraph tree in the ordinary sense rooted at e. Moreover, every vertex of such a tree belongs to at most two edges of this tree. By Lemma 6, such a tree has at most k edges. In connection with this, let us fix some notation.

Definition [r-Trees and sub-r-trees]. When we say *"r-tree"*, we mean an r-uniform hypergraph tree with at most k edges in which every vertex belongs to at most two edges of this tree. By *a sub-r-tree of G* we mean an r-tree that is a subhypergraph of G.

Often, we will consider *rooted* r-trees. *The root of an r-tree* will be an edge of this r-tree, and not a vertex. Given an r-tree T with a root e_0, the *children* of e_0 are the edges adjacent to e_0, and for $e \in E(T)$ at distance d from e_0 (in T), the *children* of e are the edges adjacent to e that are at distance $d + 1$ from e_0. Naturally, the *descendants* of an $e \in E(T)$ are its children, children of children and so on. If e_1 is a descendant of e_2, then e_2 is an *ancestor* of e_1. For an r-tree T with a root e_0 and another edge e_1 of T, by $T(e_1)$ we denote the subtree of T formed by e_1 and all its descendants. We will use the following fact on sub-r-trees of r-uniform hypergraphs.

Lemma 7. *Let H be an r-uniform hypergraph with maximum edge degree at most D. Let $e_0 \in E(H)$. Then e_0 belongs to at most $(4D)^{y-1}$ sub-r-trees of H with y edges.*

Proof. Let T be a sub-r-tree of H containing e_0 with $\big|E(T)\big| = y$. Consider T as a rooted r-tree with root e_0. Order the edges of T $e_0, e_1, \ldots, e_{y-1}$ starting from e_0 using Breadth-First search. We say that T has *type* $(h_0, \ldots, h_{y-2})$ if for $i = 0, \ldots, y-2$, edge e_i has exactly h_i children. Since $h_0 + \cdots + h_{y-2} = y - 1$, the number of distinct types does not exceed the number of representation of $y-1$ as the sum of $y-1$ of ordered nonnegative summands, which equals $\binom{(y-1)+(y-1)-1}{y-2} < 4^{y-1}$. When we know the type of T, then for every edge e_i, $i \geq 1$, we know the immediate ancestor (father edge). So, we can embed a tree T of a given type, edge by edge into G. Furthermore, at each step i, $i \geq 1$, we have at most D choices for our edge among the edges of G adjacent to its father edge. Thus, e_0 belongs to at most D^{y-1} r-trees of given type with y edges. Since the number of distinct types is at most 4^{y-1}, this proves the lemma. ∎

4. Auxiliary Events

The goal of this section is to introduce the auxiliary events $\widetilde{W}(e_0, \alpha, T, l)$ that imply the "bad" events in Evolution and can be controlled. In the next section we estimate probabilities of these auxiliary events.

Definition [Events $W(e, \alpha, T, l)$]. Let $e \in E(G)$, $\alpha \in [k]$, $l \in \{1, \ldots, n\}$ and T be a sub-r-tree of G rooted at e. Then let $W(e, \alpha, T, l)$ be the event that edge e is monochromatic of color α after Stage l of Evolution, and a cause tree for this is T.

Lemma 8. *Let $e \in E(G)$, $\alpha \in [k]$, $l \in \{1, \ldots, n\}$ and T be a sub-r-tree of G rooted at e. Also, let $e_1, \ldots, e_q$ be the edges of T of the first level, i.e., the edges of T sharing a vertex with e. For $j \in [q]$, let $e \cap e_j = \{v_j\}$. Let $Q \doteq \{v_1, \ldots, v_q\}$. If $W(e, \alpha, T, l)$ occurs, then the following properties hold:*
(W1) *For every $v \in e$, $\phi(v) \in \{\alpha\} \cup \phi(Q) \subseteq \{\alpha, \alpha - 2^0, \alpha - 2^1, \ldots, \alpha - 2^{l-1}\}$ (modulo k).*
(W2) *For $j \in [q]$, $\phi(v_j) \neq \alpha$, and for distinct j and j', $\phi(v_{j'}) \neq \phi(v_j)$. In particular, if $A_j = A_j(e, \phi) = \big\{v \in e : \phi(v) = \phi(v_j)\big\}$, then all sets A_j are disjoint subsets of e.*

(W3) $I(v) \leq ls$ *for each* $v \in \bigcup_{j=1}^{q} A_j$.

Moreover, for each $j \in [q]$, *if* v_j *becomes of color* α *at Stage* l_j, *then*

(W4) $\alpha - \phi(v_j) = 2^{l_j - 1}$;

(W5) *the event* $W(e_j, \phi(v_j), T(e_j), l_j - 1)$ *occurs;*

(W6) *for every* $u \in e_j$ *with* $I(u) > (l_j - 1)s$, *we have also* $I(u) \geq I(v_j)$; *and*

(W7) *for each* $u \in A_j - v_j$, $(l_j - 1)s + 1 \leq I(u) \leq I(v_j)$.

(W8) *If* e *was already monochromatic after Stage* $l-1$, *then for each* $v \in e$, $I(v) \notin \left[s(l-1)+1, sl\right]$.

Proof. We will prove (W1)–(W8) one by one.

(W1) holds by the definition of cause trees and Lemma 3.

By the definition of cause trees, for each $j \in [q]$, v_j is the last vertex of color $\phi(v_j)$ that changed its color to α. This implies both statements of (W2).

Since each $v \in \bigcup_{j=1}^{q} A_j$ has changed its color by Stage l, by condition (C1) in the definition of Evolution, (W3) follows.

(W4) also follows from the definition of Evolution.

If e_j were not monochromatic of color $\phi(v_j)$ after Stage $l_j - 1$, then v_j would not obtain color α blaming e_j. This yields (W5).

If some $u \in e_j$ would have $(l_j - 1)s < I(u) < I(v_j)$, then by the definition of Evolution, it would mean that u did not change its color before Stage l_j. Recall that by (W5) edge e_j is monochromatic after Stage $l_j - 1$. Thus if Condition (C2) holds at the moment $I(u)$, then vertex u should change its color at this moment, i.e. earlier than v_j did. If not, this means that some other vertex of e_j has already changed its color. In both cases, v_j would not blame e_j. This contradiction proves (W6).

Now (W7) follows from the facts that all vertices in A_j must change their colors in Stage l_j (in order to change it from $\phi(v_j)$ to α) and that v_j is the last vertex in A_j that changes its color.

If e were monochromatic after Stage $l-1$, and $I(v) \in \left[s(l-1)+1, sl\right]$ for some $v \in e$, then v would change its color, and so $W(e, \alpha, T, l)$ would not happen. This proves (W8). ■

Unfortunately, events $W(e, \alpha, T, l)$ and $W(e', \alpha', T', l')$ can be dependent even if $V(T')$ is disjoint from $V(T)$. For example, some edge $e_0 \in E(G)$ may intersect both $V(T)$ and $V(T')$. In this case, event $W(e, \alpha, T, l)$ may prompt a vertex $v \in e_0 \cap V(T)$ to change its color. This in turn, can make

e_0 monochromatic and so prompt a vertex $v' \in e_0 \cap V(T')$ to change its color, which may affect $W(e', \alpha', T', l')$.

Therefore, for each $e_0 \in E(G)$, each sub-r-tree T of G with root e_0 and $\big|E(T)\big| \leq k$, and each color α, we will introduce the auxiliary event $\widetilde{W}(e_0, \alpha, T, l)$ that contains the event $W(e_0, \alpha, T, l)$, and in addition essentially possesses properties (W1)–(W8) above, but does not depend on the values of $\phi(u)$ and $I(u)$ for all $u \notin V(T)$. This will imply that each $\widetilde{W}(e, \alpha, T, l)$ is independent of all $\widetilde{W}(e'_0, \alpha', T', l')$ with $V(T') \cap V(T) = \emptyset$.

We define these events by induction on the number of edges in T.

If $E(T) = \{e_0\}$, then the event $\widetilde{W}(e_0, \alpha, T, l)$ means that all of the following holds

(i) $\phi(e_0)$ is monochromatic of color α,

(ii) $\big|M(e_0, n)\big| < z$, and

(iii) $I(v) > ls$ for every $v \in e_0$.

Suppose that the event $\widetilde{W}(e_0, \alpha, T, l)$ is defined for all parameters e_0, α, T, l such that $\big|E(T)\big| < y$. Let $e_0 \in E(G)$, $\alpha \in [k]$, and T be any sub-r-tree T of G with root e_0 and y edges. Let $e_1, \ldots, e_q$ be the edges of T sharing a vertex with e. For $j \in [q]$, let $e \cap e_j = \{v_j\}$. Let $Q \doteq \{v_1, \ldots, v_q\}$. We say that $\widetilde{W}(e_0, \alpha, T, l)$ occurs, if either $\big|M(e, n)\big| \geq z$ for at least one $e \in E(T)$ or all of the following holds:

($\widetilde{W}1$) *For every* $v \in e$, $\phi(v) \in \{\alpha\} \cup \phi(Q) \subseteq \big\{\alpha, \alpha - 2^0, \alpha - 2^1, \ldots, \alpha - 2^{l-1}\big\}$ *(modulo* k*).*

($\widetilde{W}2$) *For* $j \in [q]$, $\phi(v_j) \neq \alpha$, *and for distinct* j *and* j', $\phi(v_{j'}) \neq \phi(v_j)$. *In particular, if* $A_j = A_j(e, \phi) = \big\{v \in e : \phi(v) = \phi(v_j)\big\}$, *then all sets* A_j *are disjoint.*

($\widetilde{W}3$) $I(v) \leq ls$ *for each* $v \in \bigcup_{j=1}^{q} A_j$.

Moreover, for each $j \in [q]$, *if* $(l_j - 1)s + 1 \leq I(v_j) \leq sl_j$, *then*

($\widetilde{W}4$) $\alpha - \phi(v_j) = 2^{l_j - 1}$;

($\widetilde{W}5$) *event* $\widetilde{W}(e_j, \phi(v_j), T(e_j), l_j - 1)$ *occurs;*

($\widetilde{W}6$) *for every* $u \in e_j$ *with* $I(u) > (l_j - 1)s$, *we have also* $I(u) \geq I(v_j)$, and

($\widetilde{W}7$) *for each* $u \in A_j - v_j$, $(l_j - 1)s + 1 \leq I(u) \leq I(v_j)$.

($\widetilde{W}8$) *If event* $\widetilde{W}(e_0, \alpha, T, l - 1)$ *occurs, then for each* $v \in e_0$, $I(v) \notin \big[s(l-1) + 1, sl\big]$.

The following two lemmas justify the introduction of the events $\widetilde{W}(e, \alpha, T, l)$.

Lemma 9. *Let $e_0 \in E(G)$, $\alpha \in [k]$, $l \in \{0, \dots, n\}$ and T be a sub-r-tree of G with root e_0. If the event $W(e_0, \alpha, T, l)$ occurs, then the event $\widetilde{W}(e_0, \alpha, T, l)$ also occurs.*

Proof. Suppose that for some values of the parameters e_0, T, l, and α, $W(e_0, \alpha, T, l)$ occurs but $\widetilde{W}(e_0, \alpha, T, l)$ does not occur. We may assume that l is minimal with this property, i.e., that for all quadruples (e_0', α', T', l') with $l' < l$ the lemma holds.

Let us check which of the properties in the definition of $\widetilde{W}(e_0, \alpha, T, l)$ may fail. Since $(\widetilde{W}1)$ and $(\widetilde{W}2)$ coincide with (W1) and (W2), respectively, they hold. For the same reason, properties $(\widetilde{W}4)$, $(\widetilde{W}6)$, and $(\widetilde{W}7)$ hold. Property $(\widetilde{W}5)$ follows from (W5) and the minimality of our counterexample. Property $(\widetilde{W}3)$ follows from the fact that otherwise, by the definition of Evolution (Condition (C1)), some vertex in $\bigcup_{j=1}^{q} A_j$ would not change its color to α.

Assume finally that $(\widetilde{W}8)$ does not hold, in other words, that $\widetilde{W}(e_0, \alpha, T, l-1)$ occurs, and for some $v \in e_0$, $I(v) \in [s(l-1)+1, sl]$. By (W8), this implies that $W(e_0, \alpha, T, l-1)$ does not occur, i.e., after Stage $l-1$, e_0 is not monochromatic of color α. It follows that in order e_0 to become monochromatic of color α after Stage l, we need $I(u) \in [s(l-1)+1, sl]$ for some $u \in \bigcup_{j=1}^{q} A_j$. On the other hand, by $(\widetilde{W}3)$ for the event $\widetilde{W}(e_0, \alpha, T, l-1)$, $I(u) \le (l-1)s$ for each $u \in \bigcup_{j=1}^{q} A_j$. This contradiction finishes the proof of the lemma. ■

Lemma 10. *Let $e_0 \in E(G)$, $\alpha_0 \in [k]$, $l_0 \in \{0, \dots, n\}$, and T_0 be a sub-r-tree of G with root e_0. Then $\widetilde{W}(e_0, \alpha_0, T_0, l_0)$ is independent of all events $\widetilde{W}(e, \alpha, T, l)$ such that $V(T) \cap V(T_0) = \emptyset$.*

Proof. By definition, the events $\widetilde{W}\big(e_0, \alpha_0, T_0, l_0, \psi(e_0)\big)$ are completely defined when we know the values of $\phi(v)$ and $I(v)$ for all $v \in V(T_0)$. This yields the lemma. ■

5. Probabilities of the Auxiliary Events

The whole proof is based on the following lemma.

Lemma 11. *Let* $D := \varepsilon k^r \left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}}$ *and* G *be an* r*-uniform hypergraph with maximum edge degree at most* D*. Let* $e \in E(G)$*,* $\alpha \in [k]$*, and* $0 \le l \le n$*. Let* T *be a rooted sub-*r*-tree of* G *with root* e*. If* T *has* y *edges, then*

$$\mathbf{Pr}\left(\widetilde{W}(e,\alpha,T,l)\right) \le \varepsilon D^{-y}\left(\frac{r}{\ln r}\right)^{\frac{n-l}{n+1}}.$$

Proof. We start from proving the (simple) auxiliary inequality

$$D^k \varepsilon^{0.5r} \le \frac{\varepsilon}{100k}. \tag{8}$$

By definition and (3),

$$D = \varepsilon k^r \left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}} \le \frac{\varepsilon}{\ln r} k^r r \le \frac{\varepsilon}{2\varepsilon^{-2(n+1)}} k^{2r} \le \frac{\varepsilon}{2}\varepsilon^4 k^{2r}.$$

Since $\varepsilon^4 = e^{-16k^2} < \frac{1}{100k}$, (8) would follow from the fact that $\left(k^{2r}\right)^k \varepsilon^{0.5r} \le 1$, which is equivalent to

$$\left(k^{2r}\right)^k \le \left(e^{4k^2}\right)^{0.5r}.$$

The last inequality is obvious.

To prove the lemma, we use induction on l. Consider first $l = 0$. If $\widetilde{W}(e,\alpha,T,0)$ occurs, then by $(\widetilde{W}1)$, $\phi(v) = \alpha$ for each $v \in e$. Thus, in this case

$$\mathbf{Pr}\left(\widetilde{W}(e,\alpha,T,0)\right) = k^{-r} = \frac{\varepsilon}{D}\left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}}.$$

This proves the case $l = 0$.

Now, suppose that the lemma holds for every $l' < l$. Consider the event $\widetilde{W}(e,\alpha,T,l)$ for some $e \in E(G)$, an r-tree T with y edges rooted at e, and $\alpha \in [k]$.

Let $X(T)$ denote the event that for at least one $e' \in E(T)$, the set $M(e',n) = \left\{v \in e' : I(v) \le sn\right\}$ (cf. (6)) has at least $z = \lfloor 4k^2 r/\ln r \rfloor$ elements. More shortly,

$$X(T) = \bigcup_{e' \in E(T)} \left\{\left|M(e',n)\right| \ge z\right\}. \tag{9}$$

Let $\overline{X(T)}$ denote the complement of $X(T)$. Suppose that the event $\widetilde{W}(e,\alpha,T,l) \cap \overline{X(T)}$ occurs.

When a sub-r-tree T with root e is given, it automatically defines the edges $e_1, \dots, e_q$ of T that share a vertex with e. It also defines for every $j \in [q]$, the vertex $\{v_j\} = e \cap e_j$ and the number y_j of edges in $T(e_j)$. Below, when T is given, we will denote $Q = Q(T, e) \doteq \{v_1, \dots, v_q\}$. Every outcome of Evolution such that $\widetilde{W}(e, \alpha, T, l) \cap \overline{X(T)}$ occurs defines the vector $(l_1, \dots, l_q)$ such that $s(l_j - 1) < I(v_j) \le s l_j$ for each $j \in [q]$. By $(\widetilde{W}3)$, for each $j \in [q]$, $l_j \in [l]$. By $(\widetilde{W}4)$ and $(\widetilde{W}2)$, all l_j are distinct.

Let $\Theta_0 = \Theta_0(q, l)$ be the set of vectors $(l_1, \dots, l_q)$ such that

$$\text{all } l_1, \dots, l_q \text{ are distinct and belong to } [l]. \tag{10}$$

By the previous paragraph,

$$\widetilde{W}(e, \alpha, T, l) \cap \overline{X(T)} = \widetilde{W}(e, \alpha, T, l) \cap \overline{X(T)} \cap \big\{ (l_1, \dots, l_q) \in \Theta_0 \big\}. \tag{11}$$

Let $\Theta_1(q, l) = \Theta_0(q, l-1)$, i.e. the set of $(l_1, \dots, l_q) \in \Theta_0$ such that $l_j \le l-1$ for all $j \in [q]$. Let $\Theta_2 = \Theta_2(q, l) = \Theta_0(q, l) - \Theta_1(q, l)$. For $i = 1, 2$, let

$$F_i(e, \alpha, T, l) = \widetilde{W}(e, \alpha, T, l) \cap \overline{X(T)} \cap \big\{ (l_1, \dots, l_q) \in \Theta_i \big\}.$$

By (11),

$$\widetilde{W}(e, \alpha, T, l) \subseteq X(T) \cup F_1(e, \alpha, T, l) \cup F_2(e, \alpha, T, l). \tag{12}$$

Our goal is to prove that for $i = 1, 2$,

$$\mathbf{Pr}\big(F_i(e, \alpha, T, l)\big) \le 0.4 \varepsilon D^{-y} \Big(\frac{r}{\ln r}\Big)^{\frac{n-l}{n+1}}. \tag{13}$$

Since by Lemma 2 and (8), $\mathbf{Pr}\big(X(T)\big) \le k\varepsilon^{0.5r} < 0.1\varepsilon D^{-y}$, (12) and (13) will imply the lemma.

Observe that the condition "$l_j \le l-1$ for all $j \in [q]$" in the definition of $\Theta_1(\alpha, l)$ implies that if $\widetilde{W}(e, \alpha, T, l)$ occurs, then all conditions $(\widetilde{W}1)$–$(\widetilde{W}8)$ are satisfied for the event $\widetilde{W}(e, \alpha, T, l-1)$. Thus,

$$F_1(e, \alpha, T, l) \subseteq \widetilde{W}(e, \alpha, T, l-1). \tag{14}$$

By the induction assumption,

$$\mathbf{Pr}\big(\widetilde{W}(e, \alpha, T, l-1)\big) \le \varepsilon D^{-y} \Big(\frac{r}{\ln r}\Big)^{(n-l+1)/(n+1)}. \tag{15}$$

Let $Z(e,l)$ be the event that $I(v) > sl$ for each $v \in e - M(e,l-1)$. If $\widetilde{W}(e,\alpha,T,l) \cap \overline{X(T)}$ holds, then by $(\widetilde{W}8)$, $Z(e,l)$ occurs. Thus

$$F_1(e,\alpha,T,l) \subseteq Z(e,l). \tag{16}$$

Since we already know whether the event $\widetilde{W}(e,\alpha,T,l-1)$ occurs or not after Step $s(l-1)$, for each $v \in e - M(e,l-1)$,

$$\begin{aligned}\mathbf{Pr}\left(\{I(v) > sl \mid \widetilde{W}(e,\alpha,T,l-1)\}\right) &\le \frac{\mathbf{Pr}\left(I(v) > sl\right)}{\mathbf{Pr}\left(I(v) > s(l-1)\right)} \\ &= \frac{1-lps}{1-(l-1)ps}.\end{aligned}$$

Since all random variables $I(v)$ are mutually independent,

$$\begin{aligned}\mathbf{Pr}\left(\{Z(e,l) \mid \widetilde{W}(e,\alpha,T,l-1)\}\right) &\le \left(\frac{1-lps}{1-(l-1)ps}\right)^{r-|M(e,l-1)|} \\ &\le (1-ps)^{r-|M(e,l-1)|}.\end{aligned} \tag{17}$$

Therefore,

$$\begin{aligned}&\mathbf{Pr}\left(\{Z(e,l) \mid \widetilde{W}(e,\alpha,T,l-1)\}\right) \\ &\le \sum_{M \subset e} \mathbf{Pr}\{M = M(e,l-1)\}(1-ps)^{r-|M|}.\end{aligned}$$

By Lemma 2, $\mathbf{Pr}\left(|M(e,l-1)| \ge z\right) \le \varepsilon^{0.5r}$. Hence

$$\begin{aligned}&\sum_{M \subseteq e} \mathbf{Pr}\{M = M(e,l-1)\}(1-ps)^{r-|M|} \\ &\le \varepsilon^{0.5r} + \sum_{M \subseteq e:\,|M|<z} \mathbf{Pr}\{M = M(e,l-1)\}(1-ps)^{r-|M|} \\ &\le \varepsilon^{0.5r} + (1-ps)^{r-z} \le \varepsilon^{0.5r} + \exp\left\{-psr\left(1-\frac{4k^2}{\ln r}\right)\right\}.\end{aligned}$$

Since $ps \ge \frac{\ln r}{(n+1)r} - p$, by the definition of p and s,

$$\begin{aligned}\exp\left\{-psr\left(1-\frac{4k^2}{\ln r}\right)\right\} &\le \exp\left\{-\left(\frac{\ln r}{n+1} - pr\right)\left(1-\frac{4k^2}{\ln r}\right)\right\} \\ &\le \exp\left\{-\frac{\ln r}{n+1} + \frac{4k^2}{n+1} + pr\right\}.\end{aligned}$$

Recall that $4k^2 = -\ln\varepsilon$. Since $p < 2^{-k^r r}$, $pr < \frac{4k^2}{n+1}$ and hence

$$\exp\left\{-\frac{\ln r}{n+1} + \frac{4k^2}{n+1} + pr\right\} \le r^{\frac{-1}{n+1}} e^{8k^2/(n+1)} \le \frac{1}{\varepsilon} r^{\frac{-1}{n+1}}.$$

By (3), $\ln r \ge 2\varepsilon^{-2(n+1)}$ and by (4), $\varepsilon = \exp\{-4k^2\}$. So,

$$\frac{1}{\varepsilon} r^{\frac{-1}{n+1}} \le \frac{1}{\varepsilon}\left(\frac{r}{\ln r}\right)^{\frac{-1}{n+1}} \varepsilon^2 = \left(\frac{r}{\ln r}\right)^{\frac{-1}{n+1}} \exp\{-4k^2\} < 0.1\left(\frac{r}{\ln r}\right)^{\frac{-1}{n+1}}. \tag{18}$$

By (14) and (16),

$$\mathbf{Pr}\left(F_1(e,\alpha,T,l)\right) \le \mathbf{Pr}\left(Z(e,l) \cap \widetilde{W}(e,\alpha,T,l-1)\right).$$

Thus by (15) and (18),

$$\begin{aligned}
&\mathbf{Pr}\left(F_1(e,\alpha,T,l)\right)\\
&\le \mathbf{Pr}\left(\left\{Z(e,l) \mid \widetilde{W}(e,\alpha,T,l-1)\right\}\right)\mathbf{Pr}\left(\widetilde{W}(e,\alpha,T,l-1)\right)\\
&\le \left(\varepsilon^{0.5r} + 0.1\left(\frac{r}{\ln r}\right)^{\frac{-1}{n+1}}\right)\mathbf{Pr}\left(\widetilde{W}(e,\alpha,T,l-1)\right)\\
&\le \varepsilon^{0.5r} + 0.1\varepsilon D^{-y}\left(\frac{r}{\ln r}\right)^{\frac{n-l}{n+1}}.
\end{aligned}$$

Since by (8), $\varepsilon^{0.5r-1} < 0.03D^{-k} \le 0.03D^{-y}$, this implies (13) for $i = 1$.

Now we will prove (13) for $i = 2$. Suppose that $F_2(e,\alpha,T,l)$ occurs.

• Then, by definition, $\vec{\ell} = (l_1,\dots,l_q) \in \Theta_2 = \Theta_2(q,l)$. In particular, there exists $j^* \in [q]$ such that $l_{j^*} = l$.

• By $(\widetilde{W}5)$, for every $j \in [q]$, the event $\widetilde{W}(e_j, \alpha - 2^{l_j-1}, T(e_j), l_j - 1)$ occurs.

• Also, for every $j \in [q]$, there exists $h_j \in [s]$ such that $I(v_j) = s(l_j - 1) + h_j$.

• By $(\widetilde{W}6)$, for every $j \in [q]$, for every $u \in e_j$ with $I(u) > (l_j - 1)s$, we have also $I(u) \ge (l_j - 1)s + h_j$.

• By $(\widetilde{W}1)$ and $(\widetilde{W}2)$, coloring $\phi(e)$ belongs to the set $\Psi(l_1,\dots,l_q)$ of colorings ψ of e such that the following two properties hold:

(P1) $\psi(v_j) = \alpha - 2^{l_j-1}$ for all $j \in [q]$.

(P2) $\psi(v) \in \{\psi(v_1), \ldots, \psi(v_q), \alpha\}$ for all $v \in e$.

For $j \in [q]$, let $A_j = \{v \in e : \phi(v) = \phi(v_j)\}$ and $a_j = |A_j| - 1$.

• By $(\widetilde{W}7)$, for every $j \in [q]$, for each $u \in A_j - v_j$, $(l_j - 1)s + 1 \leq I(u) \leq (l_j - 1)s + h_j$.

Thus, in order for $F_2(e, \alpha, T, l)$ to occur, all of the following events should occur:

(R1) There exists a vector $\vec{\ell} = (l_1, \ldots, l_q) \in \Theta_2 = \Theta_2(q, l)$ such that for every $j \in [q]$, the event $R_1(j, l_j) := \widetilde{W}(e_j, \alpha - 2^{l_j - 1}, T(e_j), l_j - 1)$ occurs. Let $R_1(\vec{\ell}) := \bigcap_{j=1}^{q} R_1(j, l_j)$.

(R2) For this vector $\vec{\ell} = (l_1, \ldots, l_q)$, there exists a vector $\vec{h} = (h_1, \ldots, h_q) \in [s]^q$ such that for every $j \in [q]$, the event $R_2(j, l_j, h_j) := \{I(v_j) = s(l_j - 1) + h_j\}$ occurs. Let $R_2(\vec{\ell}, \vec{h}) := \bigcap_{j=1}^{q} R_2(j, l_j, h_j)$.

(R3) For these vectors $\vec{\ell} = (l_1, \ldots, l_q)$ and $\vec{h} = (h_1, \ldots, h_q)$, for every $j \in [q]$ and every $u \in e_j$, the event

$$R_3(u, j, l_j, h_j) := \{I(u) \notin [s(l_j - 1) + 1, s(l_j - 1) + h_j - 1]\}$$

occurs. Let $R_3(j, l_j, h_j) := \bigcap_{u \in e_j - v_j} R_3(u, j, l_j, h_j)$ and $R_3(\vec{\ell}, \vec{h}) := \bigcap_{j=1}^{q} R_3(j, l_j, h_j)$.

(R4) For this vector $\vec{\ell} = (l_1, \ldots, l_q)$, there exists a coloring $\psi \in \Psi(\vec{\ell})$ such that the event $R_4(\phi, \vec{\ell}) := \{\phi(e) = \psi\}$ occurs.

(R5) For these $\vec{\ell}$, $\vec{h}$, and ψ, if for $j \in [q]$, we define $A_j = A_j(\psi) := \{v \in e : \psi(v) = \psi(v_j)\}$, then for every $j \in [q]$ and $u \in A_j$, the event

$$R_5(\psi, u, j, l_j, h_j) := \{(l_j - 1)s + 1 \leq I(u) \leq (l_j - 1)s + h_j\}$$

occurs. Let $R_5(\psi, \vec{\ell}, \vec{h}) := \bigcap_{j=1}^{q} \bigcap_{u \in A_j - v_j} R_5(\psi, u, j, l_j, h_j)$.

We conclude that

$$F_2(e, \alpha, T, l) \subseteq \bigcup_{\vec{\ell} \in \Theta_2} R_1(\vec{\ell}) \cap \left[\bigcup_{\vec{h} \in [s]^q} \Big(R_2(\vec{\ell}, \vec{h}) \cap R_3(\vec{\ell}, \vec{h}) \right.$$

$$\left. \cap \bigcup_{\psi \in \Psi(\vec{\ell})} \big(R_4(\psi, \vec{\ell}) \cap R_5(\psi, \vec{\ell}, \vec{h}) \big) \Big) \right].$$

So, we estimate

$$\begin{aligned}(19)\quad &\mathbf{Pr}\left(F_2(e,\alpha,T,l)\right)\\ &\le \sum_{\vec{\ell}\in\Theta_2}\mathbf{Pr}\left(R_1(\vec{\ell})\right)\sum_{\vec{h}\in[s]^q}\mathbf{Pr}\left(\left\{R_2(\vec{\ell},\vec{h})\cap R_3(\vec{\ell},\vec{h})\mid R_1(\vec{\ell})\right\}\right)\\ &\quad\times\sum_{\psi\in\Psi(\vec{\ell})}\mathbf{Pr}\left(\left\{R_4(\psi,\vec{\ell})\mid R_1(\vec{\ell})\cap R_2(\vec{\ell},\vec{h})\cap R_3(\vec{\ell},\vec{h})\right\}\right)\\ &\quad\times\mathbf{Pr}\left(\left\{R_5(\psi,\vec{\ell},\vec{h})\mid R_1(\vec{\ell})\cap R_2(\vec{\ell},\vec{h})\cap R_3(\vec{\ell},\vec{h})\cap R_4(\psi,\vec{\ell})\right\}\right).\end{aligned}$$

We now will gradually evaluate and simplify the expression in (19). We start from $R_1(\vec{\ell})$. Since the vertex sets of $T(e_j)$ for distinct j are disjoint and by Lemma 10, for every j the event $R_1(j,l_j)=\widetilde{W}\left(e_j,\alpha-2^{l_j-1},T(e_j),l_j-1\right)$ depends only on the values of $I(v)$ and $\phi(v)$ for $v\in V\left(T(e_j)\right)$,

$$\mathbf{Pr}\left(R_1(\vec{\ell})\right)=\prod_{j=1}^{q}\mathbf{Pr}\left(R_1(j,l_j)\right).$$

If $T(e_j)$ has y_j edges, then by the induction assumption,

$$\mathbf{Pr}\left(R_1(j,l_j)\right)\le\varepsilon D^{-y_j}\left(\frac{r}{\ln r}\right)^{(n-l_j+1)/(n+1)}.$$

Thus,

$$(20)\qquad \mathbf{Pr}\left(R_1(\vec{\ell})\right)\le\prod_{j=1}^{q}\varepsilon D^{-y_j}\left(\frac{r}{\ln r}\right)^{(n-l_j+1)/(n+1)}.$$

Now, consider the events $R_3(u,j,l_j,h_j)$. Since all $I(v)$ are independent and all trees T_j are vertex-disjoint, $R_3(u,j,l_j,h_j)$ does not depend on $R_1(j',l_{j'})$ for all $j'\neq j$. Furthermore, if $I(u)>s(l_j-1)$, then $\widetilde{W}\left(e_j,\alpha-2^{l_j-1},T(e_j),l_j-1\right)$ does not depend on the particular value of $I(u)$ in this range. So,

$$\mathbf{Pr}\left(\left\{R_3(u,j,l_j,h_j)\mid R_1(\vec{\ell})\right\}\right)\le\frac{1-p(l_j-1)s-p(h_j-1)}{1-p(l_j-1)s}\le 1-p(h_j-1).$$

By the independence of $I(u)$ for distinct u, similarly to (17) and the argument following (17), we have

$$\begin{aligned}
&\mathbf{Pr}\big(\{R_3(j,l_j,h_j) \mid R_1(\vec{\ell})\}\big)\\
&\le \sum_{M\subseteq e_j} \mathbf{Pr}\{M = M(e_j,l_j-1)\}\big(1-p(h_j-1)\big)^{r-|M|-1} \le \varepsilon^{0.5r}\\
&\quad + \sum_{M\subseteq e\,:\,|M|<z} \mathbf{Pr}\{M = M(h_j-1)\}\big(1-p(h_j-1)\big)^{r-|M|-1}\\
&\le \varepsilon^{0.5r} + \big(1-p(h_j-1)\big)^{r-z}.
\end{aligned}$$

Since $\varepsilon^{0.5r} \le 0.1(1-ps)^r \le 0.1\big(1-p(h_j-1)\big)^{r-z}$, again by the independence of $I(u)$ for distinct u, we conclude that

$$\mathbf{Pr}\big(\{R_3(\vec{\ell},\vec{h}) \mid R_1(\vec{\ell})\}\big) \le \prod_{j=1}^{q} 1.1\big(1-p(h_j-1)\big)^{r-z}. \tag{21}$$

Now, we consider events $R_2(j,l_j,h_j) = \{I(v_j) = s(l_j-1)+h_j\}$. Let $j\in[q]$. Similarly to $R_3(u,j,l_j,h_j)$, $R_2(j,l_j,h_j)$ does not depend on $R_1(j',l_{j'})$ for all $j'\neq j$, and

$$\mathbf{Pr}\big(\{R_2(j,l_j,h_j) \mid R_1(\vec{\ell})\}\big) \le \frac{p}{1-ps(l_j-1)} \le p(1+\frac{\ln r}{r}).$$

Moreover, again by the independence of $I(u)$ for distinct u, $R_2(j,l_j,h_j)$ is independent of all other $R_2(j',l_{j'},h_{j'})$ and of $R_3(\vec{\ell},\vec{h})$. Thus, since $q\le k \ll \ln r$,

$$\mathbf{Pr}\big(\{R_2(\vec{\ell},\vec{h}) \mid R_1(\vec{\ell})\cap R_3(\vec{\ell},\vec{h})\}\big) \le p^q\left(1+\frac{\ln r}{r}\right)^q \le 2p^q, \tag{22}$$

and together with (21), we obtain

$$\mathbf{Pr}\big(\{R_2(\vec{\ell},\vec{h})\cap R_3(\vec{\ell},\vec{h}) \mid R_1(\vec{\ell})\}\big) \le 2p^q\prod_{j=1}^{q} 1.1\big(1-p(h_j-1)\big)^{r-z}. \tag{23}$$

Now for given $\vec{\ell},\vec{h}$, and $\psi\in\Psi(\vec{\ell})$, we evaluate

$$\widetilde{P}_4(\psi,\vec{\ell},\vec{h}) := \mathbf{Pr}\big(\{R_4(\psi,\vec{\ell}) \mid R_1(\vec{\ell})\cap R_2(\vec{\ell},\vec{h})\cap R_3(\vec{\ell},\vec{h})\}\big)$$

Observe that each event $R_1(\vec{\ell}) \cap R_2(\vec{\ell}, \vec{h}) \cap R_3(\vec{\ell}, \vec{h})$ already fixes the colors of $v_1, \ldots, v_q$ in ϕ, but all other vertices of e are "free". Since for each $v \in e$, the value of $\phi(v)$ is chosen independently of all other vertices, for each $\vec{\ell}$, $\vec{h}$, and $\psi \in \Psi(\vec{\ell})$, we have

$$\widetilde{P}_4(\psi, \vec{\ell}, \vec{h}) \leq k^{q-r}. \tag{24}$$

Finally, for given $\vec{\ell}, \vec{h}$, and $\psi \in \Psi(\vec{\ell})$, consider the event

$$\widetilde{R}_5(\psi, \vec{\ell}, \vec{h}) := \left\{ R_5(\psi, \vec{\ell}, \vec{h}) \mid R_1(\vec{\ell}) \cap R_2(\vec{\ell}, \vec{h}) \cap R_3(\vec{\ell}, \vec{h}) \cap R_4(\psi, \vec{\ell}) \right\}.$$

For $j \in [q]$, let $A_j = A_j(\psi) := \left\{ v \in e : \psi(v) = \psi(v_j) \right\}$ and $a_j = |A_j| - 1$. By definition, for every $j \in [q]$ and every $u \in A_j - v_j$, the probability of the event

$$R_5(\psi, u, j, l_j, h_j) := \left\{ (l_j - 1)s + 1 \leq I(u) \leq (l_j - 1)s + h_j \right\}$$

is at most ph_j. Since the values of $I(v)$ for $v \in e - \{v_1, \ldots, v_q\}$ do not depend on $R_1(\vec{\ell}) \cap R_2(\vec{\ell}, \vec{h}) \cap R_3(\vec{\ell}, \vec{h}) \cap R_4(\psi, \vec{\ell})$, and are independent of each other for distinct v,

$$\mathbf{Pr}\left(\widetilde{R}_5(\psi, \vec{\ell}, \vec{h})\right) \leq \prod_{j=1}^{q} (ph_j)^{a_j}.$$

Since each $\psi \in \Psi(l_1, \ldots, l_q)$ is completely defined when we choose disjoint sets $A_1 - v_1, \ldots, A_q - v_q$ in $e - \{v_1, \ldots, v_q\}$,

$$\sum_{\psi \in \Psi(\vec{\ell})} \widetilde{P}_4(\psi, \vec{\ell}, \vec{h}) \, \mathbf{Pr}\left(\widetilde{R}_5(\psi, \vec{\ell}, \vec{h})\right) \tag{25}$$

$$\leq \sum_{a_1=0}^{r} \sum_{a_2=0}^{r} \cdots \sum_{a_q=0}^{r} \binom{r}{a_1}\binom{r}{a_2} \cdots \binom{r}{a_q} k^{q-r} \prod_{j=1}^{q} (ph_j)^{a_j}.$$

Thus plugging (20), (23), and (25) into (19), we have

$$\mathbf{Pr}\left(F_2(e, \alpha, T, l)\right) \leq \sum_{\vec{\ell} \in \Theta_2} \left(\prod_{j=1}^{q} \varepsilon D^{-y_j} \left(\frac{r}{\ln r}\right)^{(n-l_j+1)/(n+1)} \right) \tag{26}$$

$$\times \sum_{\vec{h} \in [s]^q} 2p^q \left(\prod_{j=1}^{q} 1.1\big(1 - p(h_j - 1)\big)^{r-z} \right) \tag{27}$$

$$\times \sum_{a_1=0}^{r} \sum_{a_2=0}^{r} \cdots \sum_{a_q=0}^{r} \binom{r}{a_1}\binom{r}{a_2} \cdots \binom{r}{a_q} k^{q-r} (ph_j)^{a_j}.$$

We now will simplify and estimate the expression in (27). First observe that

$$\big(1-p(h_j-1)\big)^{r-z} \le (1-p)^{(r-z)(h_j-1)}.$$

Thus since $0 < p < 2^{-k^r r}$ and $h_j \le s \le \frac{\ln r}{p(n+1)r}$, we have

$$1.1\big(1-p(h_j-1)\big)^{r-z} \le 1.1(1-p)^{(r-z)(h_j-1)} \le 1.2(1-p)^{(r-z)h_j}.$$

So, the expression in (27) is at most

$$2\sum_{\vec{h}\in[s]^q} p^q\Big(\prod_{j=1}^{q} 1.2(1-p)^{(r-z)h_j}\Big)$$

$$\times \sum_{a_1=0}^{r}\sum_{a_2=0}^{r}\cdots\sum_{a_q=0}^{r}\binom{r}{a_1}\binom{r}{a_2}\cdots\binom{r}{a_q}k^{q-r}(ph_j)^{a_j}$$

$$(28)\qquad = 2k^{-r}\sum_{h_1=1}^{s}\sum_{a_1=0}^{r} 1.2p(1-p)^{(r-z)h_1}\binom{r}{a_1}k(ph_1)^{a_1}$$

$$\times\Bigg[\sum_{h_2=1}^{s}\sum_{a_2=0}^{r} 1.2p(1-p)^{(r-z)h_2}\binom{r}{a_2}k(ph_2)^{a_2}$$

$$(29)\qquad \times\Bigg[\cdots\sum_{h_q=1}^{s}\sum_{a_q=0}^{r} 1.2p(1-p)^{(r-z)h_q}\binom{r}{a_q}k(ph_q)^{a_q}\Bigg]\cdots\Bigg].$$

For $j = q, q-1, \ldots, 1$ (in this order), we can estimate

$$\sum_{h_j=1}^{s}\sum_{a_j=0}^{r} 1.2p(1-p)^{(r-z)h_j}\binom{r}{a_j}k(ph_j)^{a_j}$$

$$\le 1.2pk\sum_{h_j=1}^{s}(1-p)^{rh_j-zs}\sum_{a_j=0}^{r}\binom{r}{a_j}(ph_j)^{a_j}$$

$$\le 1.2pk(1-p)^{-zs}\sum_{h_j=1}^{s}(1-p)^{rh_j}(1+ph_j)^r$$

$$\le 1.2pke^{\frac{pzs}{1-p}}\sum_{h_j=1}^{s}(1-p)^{rh_j}(1+ph_j)^r$$

$$\leq 1.2pke^{\frac{z\ln r}{(1-p)(n+1)r}} \sum_{h_j=1}^{s} (1-p)^{rh_j}(1+p)^{rh_j}$$

$$< 1.2(ps)ke^{\frac{z\ln r}{(1-p)(n+1)r}} \leq 1.2(ps)ke^{\frac{4k^2}{(1-p)(n+1)}}.$$

Since $ps \leq \frac{\ln r}{r(n+1)}$, $n+1 \geq 2$, and $4k^2 = -\ln \varepsilon$, we have

$$1.2(ps)ke^{\frac{4k^2}{(1-p)(n+1)}} \leq 1.2\frac{\ln r}{r(n+1)}k\varepsilon^{-1/(1-p)(n+1)} < e^{-3k^2/2}\frac{k\ln r}{2r(n+1)\varepsilon}.$$

Plugging these bounds into (28)–(29), we obtain that the expression in (27) does not exceed

$$2k^{-r}\prod_{j=1}^{q}\left(e^{-3k^2/2}\frac{k\ln r}{2r(n+1)\varepsilon}\right) \leq k^{-r}\prod_{j=1}^{q}\left(e^{-3k^2/2}\frac{k\ln r}{r(n+1)\varepsilon}\right).$$

Thus, by (26)–(27)

$$\text{(30)} \qquad \mathbf{Pr}\left(F_2(e,\alpha,T,l)\right)$$

$$\leq \sum_{\vec{\ell}\in\Theta_2} k^{-r}\prod_{j=1}^{q}\left(\varepsilon D^{-y_j}\left(\frac{r}{\ln r}\right)^{\frac{n-l_j+1}{n+1}}\right)\left(e^{-3k^2/2}\frac{k\ln r}{r(n+1)\varepsilon}\right).$$

Note that

$$\text{(31)} \qquad k^{-r}D^{-y_1-\cdots-y_q} = k^{-r}D^{-y+1} \leq D^{-y}\varepsilon\left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}}.$$

Hence (recalling that $D = \varepsilon k^r\left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}}$), we may rewrite (30) as

$$\text{(32)} \quad \mathbf{Pr}\left(F_2(e,\alpha,T,l)\right)$$

$$\leq D^{-y}\varepsilon\left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}} \sum_{\vec{\ell}\in\Theta_2}\prod_{j=1}^{q}\left(\varepsilon\left(\frac{r}{\ln r}\right)^{\frac{n-l_j+1}{n+1}}\right)\left(e^{-3k^2/2}\frac{k\ln r}{r(n+1)\varepsilon}\right).$$

By the definition of Θ_2, for every $\vec{\ell} \in \Theta_2$ there is j^* such that $l_{j^*} = l$. For every $j \neq j^*$, we estimate

$$\text{(33)} \qquad \varepsilon\left(\frac{r}{\ln r}\right)^{\frac{n-l_j+1}{n+1}} e^{-3k^2/2}\frac{k\ln r}{r(n+1)\varepsilon} \leq \frac{k}{n+1}e^{-3k^2/2} < \frac{1}{n+1}.$$

For $j = j^*$ we will gain more, since $l_{j^*} = l$. Namely,

$$\varepsilon \left(\frac{r}{\ln r}\right)^{\frac{n-l+1}{n+1}} e^{-3k^2/2} \frac{k \ln r}{r(n+1)\varepsilon} \leq e^{-3k^2/2} \frac{k}{n+1} \left(\frac{\ln r}{r}\right)^{\frac{l}{n+1}}. \tag{34}$$

Plugging (33) and (34) into (32), we have

$$\begin{aligned} &\mathbf{Pr}\left(F_2(e,\alpha,T,l)\right) \\ &\leq D^{-y}\varepsilon \left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}} \sum_{\vec{\ell}\in\Theta_2} (n+1)^{-q+1} \left(e^{-3k^2/2} \frac{k}{n+1} \left(\frac{\ln r}{r}\right)^{\frac{l}{n+1}}\right). \end{aligned}$$

Since the summands in the last expression do not depend on the choice of $\vec{\ell} \in \Theta_2$ and $|\Theta_2| \leq (l+1)^q \leq (n+1)^q$, we have

$$\begin{aligned} &\mathbf{Pr}\left(F_2(e,\alpha,T,l)\right) \\ &\leq D^{-y}\varepsilon \left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}} \frac{|\Theta_2|}{(n+1)^{q-1}} \left(e^{-3k^2/2} \frac{k}{n+1} \left(\frac{\ln r}{r}\right)^{\frac{l}{n+1}}\right) \leq \\ &\leq D^{-y}\varepsilon \left(\frac{r}{\ln r}\right)^{\frac{n-l}{n+1}} e^{-3k^2/2} k \leq 0.4\varepsilon D^{-y} \left(\frac{r}{\ln r}\right)^{\frac{n-l}{n+1}}. \end{aligned}$$

This proves (13) for $i = 2$ and thus the lemma. ■

Applying Lemma 11 for $l = n$, we get the following immediate consequence.

Corollary 1. *Let $e \in E(G)$ and $\alpha \in [k]$. Let $D := \varepsilon k^r \left(\frac{r}{\ln r}\right)^{\frac{n}{n+1}}$. Let T be a rooted sub-r-tree of G with root e. If T has y edges, then*

$$\mathbf{Pr}\left(\widetilde{W}(e,\alpha,T,n)\right) \leq \varepsilon D^{-y}.$$

6. Proof of Theorem 2

Recall the following version of the Local Lemma.

Theorem 4 [1]. *Let $A_1, A_2, \dots, A_N$ be any events. Let $S_1, S_2, \dots, S_N$ be subsets of $[n]$ such that for each i, A_i is independent of the events $\{A_j : j \in ([N] - S_i)\}$. If there exist numbers $x_1, x_2, \dots, x_N \in [0,1)$ such that for all $i \in [N]$, $\mathbf{Pr}(A_i) \le x_i \prod_{j \in S_i} (1 - x_j)$, then,*

$$\mathbf{Pr}\left(\bigwedge_{i \in [N]} \overline{A_i}\right) \ge \prod_{i \in [N]} (1 - x_i) > 0.$$

Radhakrishnan and Srinivasan used it in the following form.

Lemma 12 [6]. *Let $A_1, A_2, \dots, A_N$ be any events. Let $S_1, S_2, \dots, S_N$ be subsets of $[N]$ such that for each i, A_i is independent of the events $\{A_j : j \in ([N] - S_i)\}$. If for all $i \in [N]$, $\mathbf{Pr}(A_i) < \frac{1}{2}$ and $\sum_{j \in S_i} \mathbf{Pr}(A_j) \le \frac{1}{4}$, then $\mathbf{Pr}\left(\bigwedge_{i \in [N]} \overline{A_i}\right) > 0$.*

Proof. We show that if the conditions of this lemma hold, then the conditions of Theorem 4 hold for $x_i = 2\,\mathbf{Pr}(A_i)$, $i \in [N]$. Indeed, with so defined x_i, inequality

$$\mathbf{Pr}(A_i) \le x_i \prod_{j \in S_i} (1 - x_j)$$

follows if $\prod_{j \in S_i} (1 - x_j) \ge \frac{1}{2}$ holds. Furthermore,

$$\prod_{j \in S_i} (1 - x_j) \ge 1 - \sum_{j \in S_i} x_j = 1 - 2 \sum_{j \in S_i} \mathbf{Pr}(A_j) \ge \frac{1}{2} \quad \left(\text{since } \sum_{j \in S_i} \mathbf{Pr}(A_j) \le \tfrac{1}{4}\right).$$

Hence by Theorem 4, we have the result. ■

Lemma 13. *Let $0 < \varepsilon \le 4^{-k} k^{-4}$. If $\mathbf{Pr}\left(\widetilde{W}(e, \alpha, T, n)\right) \le \varepsilon D^{-y}$ for every $\alpha \in [k]$, every sub-r-tree T of G with $y \le k$ edges and for every $e \in E(T)$, then with positive probability, none of these events occurs.*

Proof. Consider the probability space of the outcomes of Evolution. Let the events $A_1, \ldots, A_N$ be the events $\widetilde{W}(e, \alpha, T, n)$ for all $e \in E(G)$, all $\alpha \in [k]$ and all sub-r-trees T of G containing e with at most k edges. It is enough to verify that the conditions of Lemma 12 hold for our events $A_1, \ldots, A_N$. Each of the conditions $\mathbf{Pr}\left(\widetilde{W}(e, \alpha, T, n)\right) < 1/2$ immediately follows from Corollary 1. By Lemma 10, for the event $A_i = \widetilde{W}(e, \alpha, T, n)$, we can take S_i equal to the set of all events $\widetilde{W}(e', \alpha', T', n)$ such that $V(T') \cap V(T) \neq \emptyset$.

Now, fix an event $A_i = \widetilde{W}(e, \alpha, T, n)$, where T has y edges, and estimate $\sum_{j \in S_i} \mathbf{Pr}(A_j)$. Let $\widetilde{W}(e', \alpha', T', n) \in S_i$ and suppose that the size of T' is y'. Then some edge e'' of T' intersects $V(T)$ (in particular, e'' can be an edge of T, too). The number of ways to choose an edge that intersects $V(T)$ is at most $D + 1$ if $y = 1$, and is at most yD, if $y > 1$. In any case, this number is not greater than kD. By Lemma 7, G contains at most $(4D)^{y'-1}$ r-trees of size y' containing edge e''. In each of such trees, there are y' ways to choose a root, e', and k ways to choose the color α'. Since $\mathbf{Pr}\left(\widetilde{W}(e', \alpha', T', n)\right) \leq \varepsilon D^{-y'}$, it follows that

(35)

$$\sum_{j \in S_i} \mathbf{Pr}(A_j) \leq \sum_{y'=1}^{k} (kD)\,(4D)^{y'-1} y' k (\varepsilon D^{-y'}) = \sum_{y'=1}^{k} k^2 y' 4^{y'-1} \varepsilon \leq k^4 4^{k-1} \varepsilon.$$

Since $0 < \varepsilon \leq 4^{-k} k^{-4}$, the last expression in (35) is at most $1/4$. Thus we are done by Lemma 12. ∎

Now we are ready to complete the proof of the main theorem. Indeed, let G be a hypergraph satisfying the conditions of the theorem. Consider procedure Evolution. By Corollary 1, for each y-edge r-tree T, each edge $e \in E(T)$ and each $\alpha \in [k]$, $\mathbf{Pr}\left(\widetilde{W}(e, \alpha, T, n)\right) \leq \varepsilon D^{-y}$. For $k \geq 2$, we have $\varepsilon = \exp\{-4k^2\} < 4^{-k} k^{-4}$. So, by Lemma 13, with positive probability none of the events $\widetilde{W}(e, \alpha, T, n)$ occurs. It follows that in some outcome of Evolution none of the events $\widetilde{W}(e, \alpha, T, n)$ occurs. By Lemma 9, in this outcome none of the events $W(e, \alpha, T, n)$ occurs. But then the resulting k-coloring will be proper.

References

[1] N. Alon and J. Spencer, *The Probabilistic Method,* Second Edition, Wiley (2000).

[2] P. Erdős and L. Lovász, Problems and results on 3-chromatic hypergraphs and some related questions, in: *Infinite and Finite Sets,* A. Hajnal et al., editors, Colloq. Math. Soc. J. Bolyai **11**, North Holland (Amsterdam, 1975), pp. 609–627.

[3] A. V. Kostochka, Coloring uniform hypergraphs with few colors, *Random Structures and Algorithms,* **24** (2004), 1–10.

[4] A. Kostochka and M. Kumbhat, Coloring simple uniform hypergraphs with few edges, *Random Structures and Algorithms,* **35** (2009), 348–368.

[5] A. Kostochka and V. Rödl, Constructions of sparse uniform hypergraphs with high chromatic number, *Random Structures and Algorithms,* **36** (2010), 46–56.

[6] J. Radhakrishnan and A. Srinivasan, Improved bounds and algorithms for hypergraph two-coloring, *Random Structures and Algorithms,* **16** (2000), 4–32.

Alexandr V. Kostochka
University of Illinois,
Urbana, IL, 61801
and
Institute of Mathematics,
Novosibirsk,
630090, Russia
e-mail: kostochk@math.uiuc.edu

Mohit Kumbhat
Department of Mathematics,
University of Illinois,
Urbana, IL, 61801,
U.S.A.
e-mail: kumbhat2@uiuc.edu

Vojtěch Rödl
Department of Mathematics and
Computer Science,
Emory University,
Atlanta, GA 30322,
U.S.A.
e-mail: rodl@mathcs.emory.edu

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 239–286.

Extremal Graphs and Multigraphs with Two Weighted Colours

EDWARD MARCHANT* and ANDREW THOMASON

To László Lovász on his sixtieth birthday

We study the extremal properties of coloured multigraphs H, whose edge set is the union of two simple graphs H_r and H_b (thought of as red and blue edges) on the same vertex set. Let $0 \leq p \leq 1$ and let $q = 1 - p$. The extremal problem considered here, for a given fixed H, is to find the maximum weight $p|E(G_r)| + q|E(G_b)|$ of large coloured multigraphs G that do not contain H as a subgraph. In fact, motivated by applications (typically to the study of hereditary properties by means of Szemerédi's Lemma), we consider the maximum restricted to those G whose underlying graph is complete – that is, every pair of vertices is joined by at least one edge.

We describe some basic features of the extremal function in general; in particular it is shown that, for any class of forbidden graphs, the extremal function always has a finite description. We then look at some examples, including a detailed study of the case $H_r = K_{3,3}$ and $H_b = \overline{H}_r = 2K_3$.

Our approach is that of classical extremal graph theory (and, as far as general properties are concerned, mirrors the work of Brown, Erdős and Simonovits on directed graphs). The recent work of Lovász and his co-authors on graph sequences might offer an alternative approach.

1. Introduction

A 2-*coloured* multigraph H is a multigraph whose edge set is the union of two simple graphs H_r and H_b on the same vertex set $V(H)$. Let H_u be the underlying graph of H; that is, H_u is a simple graph with vertex set $V(H)$,

*Research funded by Trinity College, University of Cambridge.

two vertices being adjacent in H_u if they are adjacent in at least one of H_r and H_b. We can think of H as being a colouring of the edges of the simple graph H_u with three colours red, blue and green, an edge of H_u having one of these colours according as it is in H_r only, H_b only or in both H_r and H_b. The *order* of H, denoted by $|H|$, is defined as usual to be the number of vertices of H, namely $\big|V(H)\big|$.

We say that a (large) 2-coloured multigraph G *contains* H, written $H \subset G$, if there is an injection $f : V(H) \to V(G)$ for which $f(x)f(y) \in G_r$ whenever $xy \in H_r$ and $f(x)f(y) \in G_b$ whenever $xy \in H_b$. In terms of simple graphs, this can be thought of as H_u being a subgraph of G_u in such a way that the colours agree, except that the colour green is a "wildcard", meaning that a green edge of G_u can represent an edge of H_u of any colour – red, blue or green. In particular, a complete graph of order n whose every edge is green contains every H of order at most n. (Notice here that we used the abbreviation "graph" for "2-coloured multigraph", as we shall do often.)

We shall study the extremal properties of such 2-coloured multigraphs H. The appropriate measure, for our purposes, of the size of a 2-coloured multigraph G is obtained by giving each edge of G_r the same weight p and each edge of G_b the weight q; the size of G is then the total weight. We normalize by assuming throughout that $0 \leq p \leq 1$ and $p + q = 1$. Thus the size of a 2-coloured multigraph G is taken to be its *p-weight* $w_p(G)$, defined by

$$w_p(G) = pe(G_r) + (1-p)e(G_b) = pe(G_r) + qe(G_b)$$

where $e(G_r)$ is the number of edges in G_r and $e(G_b)$ is the number of edges in G_b. Equivalently, and this is how we shall usually think of it, $w_p(G)$ is the sum of the weights of edges in G_u, where red edges have weight p, blue edges have weight q and green edges have weight 1.

It might seem that the natural extremal problem for a fixed 2-coloured multigraph H is to find, for each p and for each n, the maximum p-weight of a large 2-coloured multigraph G on n vertices that does not contain H. More generally, given a set $\mathcal{H}$ of 2-coloured multigraphs, one would find the largest weight of a G which contains no member of $\mathcal{H}$ as a subgraph. We shall indeed look at this general question in §3.6 and in §4.

However, it it turns out to be more important for our main applications to restrict attention to those large 2-coloured multigraphs G which are *complete*, meaning that every pair of vertices in G is joined by an edge of some colour – in other words, G_u is complete. Therefore we define

$$\mathrm{kex}_p(\mathcal{H}, n) = \max\big\{ w_p(G) \,:\, |G| = n,\ G \text{ complete}, H \not\subset G \text{ for all } H \in \mathcal{H} \big\}.$$

The letter 'k' in 'kex' here is to emphasize the completeness of the multigraphs G over which we are taking the maximum. Clearly, the unrestricted maximum, taken over all (not necessarily complete) graphs G, is at least as large as $\mathrm{kex}_p(\mathcal{H}, n)$ and, as would be expected, it can sometimes be greater. Example 5.10 provides an instance of an $\mathcal{H}$ where the two maxima differ. All the same, in those cases when we are able to evaluate both maxima, they often agree.

Our interest in $\mathrm{kex}_p(\mathcal{H}, n)$ rather than the unrestricted extremal function is motivated by applications. However it is a fortuitous choice, because whilst most of the theory applies to both the restricted and the unrestricted functions, the strongest general results that we have, namely Theorems 3.23 and 3.25, are known only for $\mathrm{kex}_p(\mathcal{H}, n)$. Section §3.6 discusses why the unrestricted problem is harder.

We shall be interested mostly in single forbidden graphs H, and in this case we usually write $\mathrm{kex}_p(H, n)$ instead of $\mathrm{kex}_p\big(\{H\}, n\big)$. The general theory developed in §3 applies just as well to any class $\mathcal{H}$ (apart from the degenerate classes noted in §3.1).

It is readily verified that $\mathrm{kex}_p(\mathcal{H}, n)/\binom{n}{2}$ decreases with n (*c.f.* Katona, Nemetz and Simonovits [27]) and therefore the limit

$$\kappa_p(\mathcal{H}) = \lim_{n\to\infty} \mathrm{kex}_p(\mathcal{H}, n)\binom{n}{2}^{-1}$$

exists. Since $w_p(G) \leq \binom{n}{2}$ if $|G| = n$, we have $0 \leq \kappa_p(\mathcal{H}) \leq 1$. We shall study $\kappa_p(\mathcal{H})$ rather than $\mathrm{kex}_p(\mathcal{H}, n)$ itself; thus some of the fine detail of the extremal function is lost. Something can be said about this detail in some cases (see [31]) but in general the extremal situation can be quite complicated since, in particular, it contains the classical theory of ordinary undirected graphs (*q.v.* Bollobás [10] or Simonovits [38]), which arises when H is blue-free and $p = 1$. By contrast, the extremal density $\kappa_p(H)$ is quite well-behaved.

Complete multi-partite graphs are central to the study of extremal graph theory. The corresponding objects here are large complete 2-coloured graphs G whose vertices are partitioned into classes, such that the colour of an edge depends only on which classes its endvertices lie in. The description of which colours are used within and between the classes of G is given succinctly via the notion of a *type*, introduced in [12].

Definition 1.1. A *type* is a labelled 2-coloured complete multigraph, each of whose vertices is coloured red or blue.

Hopefully it is clear that, by saying a type τ describes a graph G, we mean that the classes of G correspond to vertices of τ; edges of G within a class have the colour of the vertex of τ, and edges between classes have the colour of the corresponding edge of τ. Associated with the p-weight $w_p(G)$ of G is a natural quantity called the *p-value* $\lambda_p(\tau)$ of τ (the definition is not needed yet but can be found in Definition 3.5).

The significance of types, and the general way in which they describe extremal functions for 2-coloured multigraphs, is laid out in §3. The principal positive results of that section are Theorem 3.23 and Theorem 3.25. The second of these theorems states that, for any class $\mathcal{H}$, there is some finite type τ, such that no member of $\mathcal{H}$ is contained in a multi-partite graph that is described by τ, and which satisfies $\kappa_p(\mathcal{H}) = \lambda_p(\tau)$. Thus the determination of $\kappa_p(\mathcal{H})$ comes down to finding a suitable type τ. Theorem 3.23 is helpful here because it states that the search can be restricted to types of a quite specific form. The situation is not quite so straightforward as for classical undirected graphs, though; in §3.5 it is shown that there might be infinitely many types that satisfy Theorem 3.25, and that there need not be a finite subclass $\mathcal{H}_0 \subset \mathcal{H}$ with $\kappa_p(\mathcal{H}_0) = \kappa_p(\mathcal{H})$.

A relationship between Ramsey games and 2-coloured multigraphs is described in §4, together with a conjecture akin to the Erdős–Stone theorem.

The results of §3 are qualitative, and so the purpose of §5 is to make a quantitative study of $\kappa_p(H)$ for a few specific H's. Some simple examples have very straightforward associated types (this is always so when $p = 1/2$) but others are less straightforward. The case when $H_r = K_{3,3}$ and $H_b = \overline{H}_r = 2K_3$ is studied in some detail; this is one of the examples that motivated our work originally, and it exhibits some interesting and unexpected (to us) features.

But it is high time that we offered some background to the whole matter, so we begin by describing the relationship of extremal 2-coloured multigraphs to Ramsey games, hereditary properties and edit distance. It is important also to discuss the work of Brown, Erdős and Simonovits, who made a study of the extremal theory of ordinary (uncoloured) multigraphs. Their work, and the additions provided by Sidorenko and by Rödl, underlies most of §3.

2. Background

There are instances in the literature of studies lying near to the framework above. One of the best known works on the extremal theory of weighted graphs is that of Füredi and Kundgen [25] who, extending an initial investigation by Bondy and Tuza [13], determined the maximum total weight that a graph G of order n can have if the edge weights are integral and the total weight of the edges within any set of k vertices is bounded. Their results were applied by De Caen and Füredi [19] and by Keevash and Sudakov [28] in their work on the extremal hypergraph problem of the Fano plane. However these studies do not involve different types of edges distinguished by colours.

One or two studies come closer to the present framework. Füredi and Simonovits [26], in their own solution of the extremal Fano plane problem, discussed the maximum weight of a 4-coloured multigraph containing no properly 3-coloured K_4, when the colours all have the same weight. Still closer to home, Diwan and Mubayi [20] considered the largest size of a 2-coloured G where the forbidden 2-coloured subgraph H has no green edges. Their measure of the size of G is actually the *minimum* of $e_r(G)$ and $e_b(G)$, though during their investigation they do look at $w_p(G)$ with $p = 1 - 1/\big(|H| - 1\big)$.

The extremal theory of 2-coloured multigraphs in precisely the form described above was (we believe) introduced by Richer [34], some of whose results are described later in §4. His motivation came from applications of Szemerédi's Regularity Lemma in certain Ramsey-type games. We don't go into detail here, but the reader familiar with the Lemma will recognize the following situation (discussed in the survey of Komlós and Simonovits [29]): when we wish to find an induced subgraph isomorphic to a fixed graph F in some very large graph, we apply Szemerédi's Lemma and obtain a reduced graph R, and then construct a 2-coloured multigraph G with $V(G) = V(R)$, the red/blue/green edges of G_u corresponding to uniform pairs of density close to one/close to zero/not close to either. Then, to find an induced F in the large graph, it is enough to find a 2-coloured subgraph $H \subset G$, where $H_r = F$ and $H_b = \overline{F}$, the complement of F. In Richer's applications the very large graph is derived from some fixed graph together with a random graph on the same vertex set; it is the edge probability p of this random graph that leads to the weighting w_p. We shall say more about these applications in §4.

2.1. The probability of hereditary properties

A graph property $\mathcal{P}$ is a class of (ordinary, simple, uncoloured) graphs closed under isomorphism. It is *hereditary* if it is closed under the taking of induced subgraphs, or, equivalently, under the removal of vertices. The relationship between hereditary properties and 2-coloured multigraphs is straightforward. Every hereditary property $\mathcal{P}$ can be described by a class of graphs $\mathcal{F}$, such that the graphs in $\mathcal{P}$ are those with no induced subgraph in $\mathcal{F}$ (just take $\mathcal{F}$ to be the graphs not in $\mathcal{P}$). Associate with each graph $F \in \mathcal{F}$ the 2-coloured multigraph H where $V(H) = V(F)$, $H_r = F$ and $H_b = \overline{F}$; thus H_u is complete and H has no green edges. Let $\mathcal{H}$ be the class of all these graphs H. We can write $\mathcal{P} = \mathrm{Forb}(\mathcal{H})$, implying that $\mathcal{P}$ is the property of forbidding members of $\mathcal{H}$ as induced subgraphs, though more precisely it is the members of $\mathcal{F}$ that are forbidden.

Prömel and Steger [33] determined the number of graphs in $\mathrm{Forb}\big(\{H\}\big)$ for a single graph H. The same was done for general hereditary properties by Alekseev [1] and by Bollobás and the second author [11], who further studied the probability that a random graph has the property $\mathcal{P}$ [12]. The latter results are stated in terms of the parameter $c_p(\mathcal{P}) = \lim_{n\to\infty} c_{n,p}(\mathcal{P})$, where $2^{-c_{n,p}\binom{n}{2}}$ is the probability that a random graph of order n with edge probability p has the property $\mathcal{P}$.

The notion of a type (Definition 1.1) was introduced in [12] in order to describe $c_p(\mathcal{P})$ (though in very slightly different language). The property $\mathcal{P}(\tau)$ consists of those graphs G for which $V(G)$ has a partition into classes V_u, $u \in V(\tau)$, so that $G[V_u]$ is complete or empty according as u is red or blue, and the bipartite graph with vertex classes $G[V_u]$ and $G[V_v]$ is complete or empty according as uv is red or blue (if uv is green then there is no condition on the bipartite graph).

It is shown in [12] that, for every hereditary property $\mathcal{P}$, and for any $\varepsilon > 0$, there is a type τ such that $\mathcal{P}(\tau) \subset \mathcal{P}$ and $c_p\big(\mathcal{P}(\tau)\big) \le c_p(\mathcal{P}) + \varepsilon$. Moreover it is shown that $c_p\big(\mathcal{P}(\tau)\big) = H_p(\tau)$, where H_p is an easily computed function (the notation H_p was used because of a relation to entropy and it is not connected to the use in this paper of, say, H_b as a graph).

In this paper we use the somewhat more natural function $\lambda_p(\tau)$ of τ (called the p-value of τ; see Definition 3.5), which is essentially the same as H_p after a change of variable; to be exact, $H_p = -(\log p + \log q)(1 - \lambda_{p'})$ where $p' = (\log q)/(\log p + \log q)$. It turns out that we can say more than was said in [12]. By Theorem 3.25 applied to $\mathcal{H}$ where $\mathcal{P} = \mathrm{Forb}\,(\mathcal{H})$,

there is a type τ with $\mathcal{P}(\tau) \subset \mathcal{P}$ for which $\lambda_p(\tau) = \kappa_p(\mathcal{H})$. This has the consequence that $c_p\big(\mathcal{P}(\tau)\big) = c_p(\mathcal{P})$. (The counterexample to the existence of such a τ, asserted in [12], is erroneous.) Moreover it is possible to give some information about the structure of τ itself, via Theorem 3.23.

The precise application of the work in this paper to the probability of hereditary properties requires a little care, and is described in a separate note [32].

2.2. Edit distance

The notion of edit distance of relevance to coloured multigraphs is that studied by Axenovich, Kézdy and Martin [7], prompted by applications in biology, and by Alon and Stav [4], motivated by applications in computer science, especially in relation to property testing.

Given a graph property $\mathcal{P}$, the *edit distance* from a graph G to $\mathcal{P}$ is

$$\mathrm{Dist}\,(G,\mathcal{P}) = \min\big\{\big|E(J)\triangle E(G)\big| \,:\, J \in \mathcal{P},\ V(J) = V(G)\big\}$$

and the edit distance from the class of all n-vertex graphs to $\mathcal{P}$ is

$$\mathrm{Dist}\,(n,\mathcal{P}) = \max\big\{\,\mathrm{Dist}\,(G,\mathcal{P}) \,:\, |G| = n\big\}.$$

In practice, only hereditary properties are considered.

The crucial discovery of Alon and Stav is that, for any hereditary property $\mathcal{P}$, random graphs $G(n,p)$ have more or less the largest distance from $\mathcal{P}$. More precisely, by developing the ideas of Alon and Shapira [3], they show the existence of some probability $p^* = p^*(\mathcal{P})$ for which $\mathrm{Dist}\,(n,\mathcal{P}) = \mathrm{Dist}\,\big(G(n,p^*),\mathcal{P}\big) + o(n^2)$ holds almost surely. But their method gives no clue as to the value of p^*. Nevertheless, the limit

$$\mathrm{e}_p(\mathcal{P}) = \lim_{n\to\infty} \mathbb{E}\,\mathrm{Dist}\,\big(G(n,p),\mathcal{P}\big)\binom{n}{2}^{-1}$$

can be shown to exist (see [4]), where $\mathbb{E}$ denotes expectation. It follows that

$$\mathrm{ed}\,(\mathcal{P}) = \lim_{n\to\infty} \mathrm{Dist}\,(n,\mathcal{P})\binom{n}{2}^{-1}$$

exists and that $\mathrm{ed}\,(\mathcal{P}) = \mathrm{e}_{p^*}(\mathcal{P})$. Since $\mathrm{e}_p(\mathcal{P}) \le \mathrm{ed}\,(\mathcal{P})$ by the definition of edit distance, it follows that $\mathrm{e}_{p^*}(\mathcal{P}) = \max_p \mathrm{e}_p(\mathcal{P})$; that is, $p^*(\mathcal{P})$ is the value of p that maximizes $\mathrm{e}_p(\mathcal{P})$.

The relationship between edit distance and 2-coloured multigraphs can now be seen. Let $\mathcal{F}$ and $\mathcal{H}$ be as in §2.1, so $\mathcal{P} = \text{Forb}(\mathcal{H})$. Now choose a value of p and a complete 2-coloured G of order n, with $H \not\subset G$ for all $H \in \mathcal{H}$, and with $w_p(G) = \text{kex}_p(\mathcal{H}, n)$. Given a random graph $G(n,p)$ on the same vertex set as G, form J by deleting those edges in $G(n,p)$ which are blue (not green) in G, and by adding those edges not already in $G(n,p)$ which are red (not green) in G. If J contains a member of $\mathcal{F}$ as an induced subgraph then G must contain a member of $\mathcal{H}$, bearing in mind that green edges of G can stand for edges of H of either colour. (Notice that the completeness of G is crucial here.) But G contains no member of $\mathcal{H}$ and so $J \in \mathcal{P}$. The expected number of edge changes needed to form J is p times the number of blue (not green) edges plus q times the number of red (not green) edges; that is, $\binom{n}{2} - w_p(G) = \binom{n}{2} - \text{kex}_p(\mathcal{H}, n)$. This gives an upper bound on the expected distance of $G(n,p)$ from $\mathcal{P}$; hence $\text{e}_p(\mathcal{P}) \leq 1 - \kappa_p(\mathcal{H})$.

The following proposition states that this inequality is tight; the assertion is implicit in [4] and explicit in Balogh and Martin [8, Theorem 11] (though stated there in different terminology).

Proposition. *Let $\mathcal{P} = \text{Forb}(\mathcal{H})$ be a hereditary property. Then, for all p, $\text{e}_p(\mathcal{P}) = 1 - \kappa_p(\mathcal{H})$ holds. In particular, $\text{ed}(\mathcal{P}) = 1 - \min_p \kappa_p(\mathcal{H})$.*

Note by Theorem 3.1 that $\kappa_p(\mathcal{H})$ is continuous in p and so the minimum exists. We sketch the reason why the proposition follows from Szemerédi's Regularity Lemma. As shown in [4], $\text{Dist}\big(G(n,p), \mathcal{P}\big)$ is concentrated near its mean, so with high probability there exists $J \in \mathcal{P}$ with $\big|E(J) \triangle E\big(G(n,p)\big)\big| = \text{e}_p(\mathcal{P})\binom{n}{2} + o(n^2)$. Take a uniform partition of J into k (a large number of) parts. The density of $G(n,p)$ between each pair of sets in this partition will be close to p. So if we form the 2-coloured multigraph G whose vertices are the k parts and whose edges are red/blue/green according as the pairs have density in J close to zero/close to one/neither of these, then the red and blue edges of G indicate places where $G(n,p)$ is far from J. In particular, with high probability $\big|E(J) \triangle E\big(G(n,p)\big)\big| \geq \big(\binom{k}{2} - w_p(G)\big)\big(\frac{n}{k}\big)^2 + o(n^2)$. But G cannot contain a member of $\mathcal{H}$ for then, assuming all pairs of the partition are uniform (see below), we could find a member of $\mathcal{F}$ induced in J. Thus $w_p(G) \leq \text{kex}_p(\mathcal{H}, k) = \kappa_p(\mathcal{H})\binom{k}{2} + o(1)$. Therefore

$$\text{e}_p(\mathcal{P})\binom{n}{2} + o(n^2) \geq \big(1 - \kappa_p(\mathcal{H})\big)\binom{k}{2}\left(\frac{n}{k}\right)^2 + o(n^2)$$

which completes the sketch.

The reader familiar with Szemerédi's Lemma will have spotted a couple of holes in the above sketch. Most obviously, we assumed all pairs of the partition were uniform, which is known to fail in general. But admitting non-uniform pairs is equivalent to admitting 2-coloured G that are not complete, in which case the inequality $w_p(G) \leq \mathrm{kex}_p(\mathcal{H}, k)$ might fail. More subtly, if $\mathcal{H}$ is infinite, it is not possible to bound in advance the number of vertices of the graph F, and its construction in J requires a level of uniformity that depends on $|F|$. Both of these difficulties can be resolved by using a development of the Regularity Lemma due to Alon, Fischer, Krivelevich and Szegedy [2], see [4, Lemma 2.7], asserting the existence of an induced subgraph $J' \subset J$ meeting each part in cn vertices and, restricted to which, every pair is uniform, and in which the degree of uniformity is bound to the number of parts. The argument can then be carried through with J' in place of J.

The proof, involving a random graph superimposed on a Szemerédi partition, is very similar to the work of Richer (see §4) except in the important detail that Richer needs only the ordinary Regularity Lemma, meaning that his investigation of extremal 2-coloured multigraphs includes non-complete G, although often in practice this makes no difference.

The edit distance of a considerable variety of properties $\mathcal{P}$ is investigated by Axenovich, Kézdy and Martin [7] and by Alon and Stav [5], often by studying the associated types. One of the outstanding properties left unresolved by [5] is that of containing no induced $K_{3,3}$. The edit distance of this property was subsequently determined by Balogh and Martin [8]. The corresponding extremal 2-coloured graph problem is particularly interesting and is examined in more detail in §5.4.

2.3. Directed graphs and multigraphs

The extremal theory for directed graphs and multigraphs was developed systematically in a series of substantial papers by Brown, Erdős and Simonovits [15, 16, 17]. An important further contribution was made by Sidorenko [36], whose theorems and counterexamples answered many unresolved questions, and another significant counterexample was supplied by Rödl and Sidorenko [35]. Brown and Simonovits have written a survey of the subject [18]. The object of all of these studies is as follows. In the multigraph situation, given a natural number m, some class of multigraphs $\mathcal{L}$ is forbidden, and it is required to determine the maximum size of multi-

graphs which contain no member of $\mathcal{L}$ and in which the edge multiplicities are bounded by m. The directed graph situation is similar; multiple edges are allowed but no more than m in each direction between any two vertices.

What is made clear in these papers is that the extremal theories for directed graphs with multiplicity at most m and for multigraphs with multiplicity at most $2m$ are very close in spirit, to the extent that results proved for one theory can be asserted for the other without real need for a second proof to be given. Nevertheless it is not possible to deduce any formal equivalence between the theories (there are of course occasional one-way implications that can be observed).

It is apparent from the methods used in these studies that they give a lot of information about the extremal properties of 2-coloured multigraphs, as we shall develop in §3. The nearest equivalent might seem to be the theory for multigraphs (and we shall not say anything more about directed graphs). Indeed, given the continuity of $\kappa_p(H)$ (Theorem 3.1) it would be enough to evaluate this function for rational values of p, in which case (by dropping the requirement that $p + q = 1$) we could assume the edge weights p and q were integers. It might then be hoped that the theory of 2-coloured multigraphs could be deduced from the ordinary theory of multigraphs with, say, $m = p + q$. However this is not the case because, crucially, in a 2-coloured multigraph we can distinguish the colour of the edges.

Particularly important in the classical undirected multigraph theory is the case $m \leq 2$, which exhibits special features which either are unknown for the cases $m \geq 3$ or which do not in fact hold. It turns out that the theory for 2-coloured multigraphs exhibits a somewhat remarkable character, in that some, but not all, of the special features found in the classical case when $m \leq 2$ hold also for 2-coloured multigraphs. In that sense the coloured theory might be thought of as "mimicking the case $m = 2\frac{1}{2}$" of the classical multigraph theory.

Let us be more specific. Turán's theorem, its extension by Erdős and Stone [24] and the developments by Erdős and Simonovits [21, 22, 23, 37] show that complete multipartite graphs are asymptotically extremal for any forbidden class $\mathcal{L}$ of simple undirected graphs, and indeed that all extremal graphs have approximately this structure (the latter phenomenon is called *stability*). For 2-coloured multigraphs, the objects that play the rôle of multipartite graphs are the classes $\mathcal{Q}(\tau)$ described by types τ (the 2-coloured version of the graph properties $\mathcal{P}(\tau)$: see Definition 3.2). An alternative way to specify the structure of the graphs in $\mathcal{Q}(\tau)$ is by means of a (symmetric)

matrix, whose rows and columns are indexed by the vertices of τ and whose entries are p, q or 1 according to the weights of the appropriate edge colours. It can be shown without too much difficulty, perhaps either by Szemerédi's Lemma or, preferably, by a simple direct argument as in §3, that for any class $\mathcal{H}$, the extremal function $\kappa_p(\mathcal{H})$ can be approximated arbitrarily closely by the values $\kappa_p\big(\mathcal{Q}(\tau)\big)$ where no $H \in \mathcal{H}$ is in $\mathcal{Q}(\tau)$.

The corresponding structures for multigraphs, in which the number of edges between groups of vertices is determined by a finite matrix, were analysed in depth by Brown, Erdős and Simonovits. They discovered several important properties of the case $m \leq 2$. In [15] they showed that there is a finite matrix that gives the *exact* limiting extremal density. Moreover in [17] they gave an algorithm to find *all* such optimal matrices (greatly simplified later by Sidorenko [36]). These optimal matrices must satisfy strong constraints [17]. Furthermore every class of forbidden multigraphs contains a finite sub-class whose extremal function is (asymptotically) equal to that of the whole class [17].

For the cases $m \geq 3$ the situation is quite different. In these cases it remains unknown whether there is an optimal matrix giving the exact extremal density. For $m \geq 3$ Sidorenko [36] has shown that there can be an infinite number of optimal matrices, killing any hope of an algorithm to find them or of proving a stability result for the extremal graphs. Finally, Rödl and Sidorenko [35] have shown that for $m \geq 4$ there need be no finite sub-class equivalent to the whole class.

Ideas used in both the cases $m \leq 2$ and $m \geq 3$ can be adapted for use here. So it is shown in Theorem 3.25 that for any class $\mathcal{H}$ of 2-coloured multigraphs there is a type τ that determines exactly the value of $\kappa_p(\mathcal{H})$. We also establish strong constraints on the nature of the types involved (Theorem 3.23), which makes the evaluation of $\kappa_p(\mathcal{H})$ a realistic possibility in many instances. On the other hand, there might be infinitely many optimal types in some cases. Likewise there need not be a finite sub-class $\mathcal{H}_0 \subset \mathcal{H}$ with $\kappa_p(\mathcal{H}_0) = \kappa_p(\mathcal{H})$.

These are the reasons why we say that 2-coloured multigraphs lie somewhere between the undirected multigraph cases $m \leq 2$ and $m \geq 3$. An explanation of sorts for this phenomenon might be that in the case $m \leq 2$ the entries in the extremal matrices can have only two values, whereas more values are possible when $m \geq 3$. In the case of complete 2-coloured multigraphs, the matrix entries take three values p, q or 1, but the value 1 is not genuinely distinct, since it denotes green edges which are really the union of a red and a blue edge. Thus we acquire some of the properties of the case

$m \leq 2$. On the other hand the above-mentioned counterexamples for the case $m \geq 3$ involve multigraphs of only three different multiplicities, and (for suitable p) their properties can be emulated here also.

2.4. Graph sequences

We remark finally that a new approach to classical extremal graph theory and related extremal problems has emerged recently via the study of graph sequences and their limits. Perhaps this approach could be fruitful in the investigation of 2-coloured multigraphs; in particular, Lovász and Szegedy [30] have re-proved the main results of [4] by this method. However, we have not pursued this idea ourselves.

We move on now, at last, to our results.

3. General Results

Following some simple observations about κ_p in §3.1, we discuss the basic properties $\mathcal{Q}(\tau)$ in §3.2. As stated already, these were introduced in [12] and are analogous to the matrix properties of [15, 16, 17]. In particular, we shall describe extensions of properties and show how all properties can be approximated by basic properties.

The remaining sub-sections contain results which are not so near the surface. In §3.3 it is shown that only certain special kinds of types need be considered in relation to extremal 2-coloured multigraphs (Theorem 3.23). The main result is Theorem 3.25 in §3.4, that for every class $\mathcal{H}$ there is a type τ that determines exactly the value of $\kappa_p(\mathcal{H})$. These results are modelled on those of Brown, Erdős and Simonovits [15, 16, 17].

The next sub-section contains examples, based on those of Sidorenko [36] and of Rödl and Sidorenko [35], to show that extremal sequences need not exhibit stability, nor need there be a finite subclass $\mathcal{H}_0 \subset \mathcal{H}$ such that $\kappa_p(\mathcal{H}_0) = \kappa_p(\mathcal{H})$.

Finally, in §3.6, we look at to what extent the results of this section carry over to the more general extremal situation for 2-coloured multigraphs, in which the underlying extremal graphs need not be complete.

3.1. Continuity and convexity

In the introduction we defined $\kappa_p(H)$ and stated that $\kappa_p(\mathcal{H})$ was defined analogously. However we must exclude classes $\mathcal{H}$ which contain both a 2-coloured graph with no red edges and another graph with no blue edges. In this case, Ramsey's theorem means that any sufficiently large 2-coloured G will contain a member of $\mathcal{H}$. We call such a class *trivial* and, from now on, unless stated otherwise, we assume that all classes $\mathcal{H}$ under consideration are non-trivial; that is, if $\mathcal{H}$ contains any monochromatic graphs then these are all of the same colour.

It is immediate from the definition that $0 \leq \kappa_p(\mathcal{H}) \leq 1$ for all p and all $\mathcal{H}$. Moreover, if no graph in $\mathcal{H}$ is monochromatic then $\kappa_0(\mathcal{H}) = \kappa_1(\mathcal{H}) = 1$, as demonstrated by monochromatic extremal graphs. On the other hand, if $\mathcal{H}$ contains a subclass $\mathcal{H}_{\mathrm{m}}$ of monochromatic graphs, say, red ones, then $\kappa_0(\mathcal{H}) = 1$ and $\kappa_1(\mathcal{H}) \leq \lim_{n\to\infty} \mathrm{ex}\,(n, \mathcal{H}_{\mathrm{m}})/\binom{n}{2}$ where $\mathrm{ex}\,(n, \mathcal{H}_{\mathrm{m}})$ is the ordinary uncoloured extremal function for $\mathcal{H}_{\mathrm{m}}$.

Let $0 \leq p_1 < p < p_2 \leq 1$, so that $p = xp_1 + (1-x)p_2$ for some $0 < x < 1$. As mentioned in the introduction, $\mathrm{kex}_p(\mathcal{H}, n)/\binom{n}{2}$ decreases with n, so for each n we can pick G^n containing no member of $\mathcal{H}$ such that $|G^n| = n$ and $w_p(G^n) \geq \kappa_p(\mathcal{H})\binom{n}{2}$. Now $w_p(G^n) = pe(G^n_r) + qe(G^n_b) = xw_{p_1}(G^n) + (1-x)w_{p_2}(G^n)$. So $x\kappa_{p_1}(\mathcal{H}) + (1-x)\kappa_{p_2}(\mathcal{H}) \geq \lim_{n\to\infty}\big(\big(xw_{p_1}(G^n) + (1-x)w_{p_2}(G^n)\big)\big)/\binom{n}{2} \geq \kappa_p(\mathcal{H})$. So $\kappa_p(\mathcal{H})$ is convex, and thus continuous for $0 < p < 1$. In fact it is continuous at the endpoints too; for clearly $\mathrm{kex}_{p'}(\mathcal{H}, n) \geq \mathrm{kex}_p(\mathcal{H}, n) - |p - p'|\binom{n}{2}$, and so $\big|\kappa_p(\mathcal{H}) - \kappa_{p'}(\mathcal{H})\big| \leq |p - p'|$ for all p, p'. Hence we have the following result (noted also in [4] and [8]).

Theorem 3.1. *The function $\kappa_p(\mathcal{H})$ is convex and continuous as a function of p.*

One useful consequence of convexity is that a local minimum is a global minimum (as needed for edit distance §2.2).

3.2. Basic facts

Recall the notion of a *type* that was given in Definition 1.1 in §1. These are the same as the *coloured regularity graphs* of [4] and [8]. Here is a further definition from [12].

Definition 3.2. Let τ be a type. The graph $m \times \tau$ is the complete 2-coloured graph G of order $m|\tau|$, whose vertex set has a partition $\{V_v : v \in V(\tau)\}$ with $|V_v| = m$, such that edges within V_v are the same colour as v and the edges between V_u and V_v are the same colour as uv, for all $u, v \in \tau$.

The *basic property* $\mathcal{Q}(\tau)$ consists of all complete 2-coloured graphs G that are subgraphs of $m \times \tau$ for some m.

Thus graphs in $\mathcal{Q}(\tau)$ are those having a vertex partition $\{V_v : v \in V(\tau)\}$ into classes of possibly different sizes, whose edge colours are the same as those in the definition of $m \times \tau$ except in the case of a green edge uv, in which case the subgraph between V_u and V_v must be complete but can be coloured in any way (and need not be monochromatic). This definition can be seen to be the 2-coloured equivalent of the property $\mathcal{P}(\tau)$ given in [12] and discussed in §2.1.

Definition 3.3. The 2-coloured graph H is said to be *τ-colourable* if it is a subgraph of some member of $\mathcal{Q}(\tau)$.

The only difference between members of $\mathcal{Q}(\tau)$ and τ-colourable graphs in general is that the latter need not be complete. The largest order of a type τ for which H is not τ-colourable is called the *colouring number* of H. This is equivalent to the parameter defined in regard to induced subgraphs by Prömel and Steger [33]; its relationship to edit distance was studied by Axenovich, Kézdy and Martin [7].

For the purposes of calculations about a type it is helpful to associate a matrix with it.

Definition 3.4. The *weight* $w_p(uv)$ of the edge uv of a 2-coloured graph is p, q or 1 according as uv is red, blue or green. The *weight* $w_p(u)$ of the vertex u of a type τ is p or q according as u is red or blue.

The p-matrix $W_p(\tau)$ of the type τ is the symmetric matrix whose rows and columns are indexed by $V(\tau)$; the u, u entry is $w_p(u)$ and the u, v entry is $w_p(uv)$.

Evidently τ determines $W_p(\tau)$, and vice-versa if $p \neq 0, \frac{1}{2}, 1$; we could, then, by and large, continue the discussion in terms of matrices rather than types, which is of course the approach of Brown, Erdős and Simonovits [15, 16, 17].

Definition 3.5. The *p-value* of the type τ is $\lambda_p(\tau) = \max_{z \in \Delta} z^t W_p(\tau) z$, where $\Delta = \Delta(|\tau|)$ is the simplex $\{z \in [0,1]^{|\tau|} : z_1 + \cdots + z_{|\tau|} = 1\}$.

Remark 3.6. This definition is prompted by the weights of members of $\mathcal{Q}(\tau)$. For suppose $G \in \mathcal{Q}(\tau)$ is such that $|G| = n$ and $\{V_v : v \in V(\tau)\}$ is a partition of $V(G)$ proving membership. The weight of G will be greatest if all edges between V_u and V_v are green whenever uv is green. Let $|V_v| = y_v n$, so that $y = (y_1, \ldots, y_{|\tau|}) \in \Delta$. Then the p-weight of G is

$$\begin{aligned} w_p(G) &= p \sum_{v \text{ is red}} \binom{y_v n}{2} + q \sum_{v \text{ is blue}} \binom{y_v n}{2} \\ &\quad + p \sum_{uv \in E(\tau_r)} y_u y_v n^2 + q \sum_{uv \in E(\tau_b)} y_u y_v n^2 \\ &= y^{\mathrm{t}} W_p(\tau) y \binom{n}{2} + O(n). \end{aligned}$$

Given any fixed $z \in \Delta$, we can choose G so that the entries of $y - z$ are $O(1/n)$, so it follows that the maximum p-weight of graphs in $\mathcal{Q}(\tau)$ is $\lambda_p(\tau)\binom{n}{2} + O(n)$.

The central point of the present paper is that we can determine $\kappa_p(\mathcal{H})$ in terms of $\lambda_p(\tau)$ for those τ such that the graphs in $\mathcal{H}$ are not τ-colourable. So it is worth stating the easy part of this relationship here, which follows from the fact that the graphs $G \in \mathcal{Q}(\tau)$ just defined contain no member of $\mathcal{H}$.

Observation 3.7. *Let $\mathcal{H}$ be a class of 2-coloured multigraphs and let τ be a type such that no graph in $\mathcal{H}$ is τ-colourable. Then $\lambda_p(\tau) \le \kappa_p(\mathcal{H})$.*

Our final definitions in this section are these.

Definition 3.8. An *extension* of the type τ is a type σ with $\tau \subset \sigma$ and $|\sigma| = |\tau| + 1$.

Definition 3.9. A *graph sequence* is a sequence $(G^n)_{n=1}^{\infty}$ of 2-coloured complete graphs with $|G^n| \ge n$. Let τ be a type. We say that (G^n) *contains* τ if for every m there is an n such that $m \times \tau$ is a subgraph of G^n.

Notice that this definition is equivalent to stating that every τ-colourable graph is a subgraph of infinitely many G^n.

Definition 3.10. The *degree* of the vertex v in the 2-coloured graph G is $d_p(v) = p\big|\{u : vu \in E(G_r)\}\big| + q\big|\{u : vu \in E(G_b)\}\big|$. The *minimum degree* is $\delta_p(G) = \min_{v \in V(G)} d_p(v)$.

The relevance of extensions to this work is provided by the next lemma.

Lemma 3.11. *Let (G^n) be a graph sequence and let*

$$\delta = \liminf_{n\to\infty} \left(\delta_p(G^n)/|G^n|\right).$$

Let τ be a type contained in (G^n). Let $x = \left(x_1, \ldots, x_{|\tau|}\right) \in \Delta$. Then (G^n) contains an extension σ of τ for which

$$\sum_{u\in V(\tau)} x_u w(uv) \geq \delta,$$

where $\{v\} = V(\sigma) - V(\tau)$.

Remark 3.12. The purpose of the vector x in this lemma is worth pointing out. It will be of use when we come to apply the lemma to specific examples in §5. Suppose we have a type τ contained in (G^n), and there is some subset of the vertices of τ such that we would like an extension of τ with the new vertex joined by green to this subset. Now the lemma provides an extension such that the weighted sum of all edge weights at v is large. To encourage the lemma to provide an extension of the kind we want, we can choose an x for which x_u is smaller for vertices u of less interest, so the contribution of the vertices of interest is amplified relative to the others.

The definition of an extension is analogous to that of an *augmentation* which plays a fundamental rôle in the work of Brown, Erdős and Simonovits[15, 16, 17]. The idea is a very general one and applies equally to directed graphs, to multigraphs and to 2-coloured multigraphs (as remarked in [18], "augmentation does not feel the orientations"). Their definitions are in terms of the corresponding p-matrices $W_p(\tau)$ and $W_p(\sigma)$. The only essential difference between the two definitions is that in the definition of an augmentation the vector x is chosen to be one for which $\lambda_p(\tau) = x^{\mathrm{t}} W_p(\tau) x$. This difference is of no significance for the proofs of the general results but we retain it for its use in §5, as just described.

Proof. If $\delta = 0$ it is enough to form σ by picking some vertex $u \in \tau$, adding a new vertex v of the same colour as u, joining it to u by an edge of that colour and joining v to every other vertex in the same way that u is joined. For then $m \times \sigma \subset 2m \times \tau$ so σ is contained in (G^n), and the final condition of the lemma is vacuous when $\delta = 0$. So we may assume $\delta > 0$.

We shall show that, for each natural number m, there is some extension σ of τ satisfying the conditions of the lemma and with $m \times \sigma \subset G^n$ for

some n. This will be enough to complete the proof, because there are only finitely many extensions of τ, and so there is some extension σ for which the inclusion $m \times \sigma \subset G^n$ occurs for infinitely many m, and so for all m (since $m \times \sigma \subset m' \times \sigma$ if $m < m'$).

So let m be fixed. Fix also $0 < \varepsilon < \delta/4$. It now suffices to find an extension σ with $m \times \sigma \subset G^n$ for some n, such that σ satisfies the weaker condition

$$\sum_{u \in V(\tau)} x_u w(uv) \geq \delta - 4\varepsilon.$$

This is because there are only finitely many extensions σ, and so some σ must satisfy the weaker condition for every $\varepsilon > 0$, and hence must satisfy the condition of the lemma.

Pick an integer M such that $M > 6m|\tau|\varepsilon^{-1}$ and also $\lfloor x_u M \rfloor - 3m > x_u M/2$ for every $u \in V(\tau)$ with $x_u \neq 0$. Now $M \times \tau \subset G^n$ holds for infinitely many n; pick such an n that satisfies $n \geq 2R(m)\varepsilon^{-1}3^{|\tau|}\binom{M}{m}^{|\tau|}$, where $R(m)$ is the two-colour Ramsey number for K_m. By increasing n if necessary we may also assume $\delta_p(G^n) \geq (\delta - \varepsilon)|G^n|$. We shall show there is an appropriate extension σ of τ with $m \times \sigma \subset G^n$.

Choose a copy of $M \times \tau \subset G^n$ and denote its vertex classes by V_u, $u \in \tau$. For each u, select a subset $X_u \subset V_u$ with $|X_u| = \lfloor x_u M \rfloor$ and let $X = \bigcup_u X_u$. Thus $M - |\tau| \leq |X| \leq M < \varepsilon n/4$.

Given a vertex $z \in V(G^n)$ and a set $Z \subset V(G)$ let $d_Z(z) = \sum_{y \in Z} w_p(zy)$. Then $\sum_{z \in V(G)} d_X(z) = \sum_{y \in X} d_p(y) \geq |X|(\delta - \varepsilon)|G^n|$. Let $Y = \{ z \notin X : d_X(z) \geq (\delta - 2\varepsilon)|X| \}$. Then

$$\begin{aligned} |X|(\delta - \varepsilon)|G^n| &\leq \sum_{z \in V(G)} d_X(z) \\ &\leq \big(|Y| + |X|\big)|X| + \big(|G^n| - |Y| - |X|\big)(\delta - 2\varepsilon)|X|. \end{aligned}$$

Thus $|Y| + |X| \geq \varepsilon|G^n|/(1 - \delta + 2\varepsilon) \geq \varepsilon n$ and so $|Y| \geq \varepsilon n/2$.

Fix for the moment a vertex $y \in Y$. We claim that, for each $u \in \tau$, there is a subset $U_u \subset V_u$ with $|U_u| = m$, so that the edges between y and U_u are all the same colour, whose weight we denote by w_u; moreover, if $x_u \neq 0$, we claim U_u may be chosen so that $w_u \geq d_{X_u}(y)/|X_u| - \varepsilon/x_u|\tau|$. If $x_u = 0$ the claim is verified simply by choosing m vertices in V_u joined to y by edges of the same colour, and this is possible since $|V_u| = M > 3m$. If $x_u \neq 0$,

observe that if the claim fails then there are less than $3m$ edges between y and X_u whose weight satisfies the bound for w_u; thus

$$\begin{aligned} d_{X_u}(y) &\leq 3m + \big(|X_u| - 3m\big)\big(d_{X_u}(y)/|X_u| - \varepsilon/x_u|\tau|\big) \\ &< 3m + d_{X_u}(y) - \big(|X_u| - 3m\big)\varepsilon/x_u|\tau| \\ &< 3m + d_{X_u}(y) - x_u M\varepsilon/2x_u|\tau| \quad \text{by the choice of } M \\ &< d_{X_u}(y) \quad \text{also by the choice of } M. \end{aligned}$$

This contradiction establishes the claim.

By the definition of Y we have $d_X(y) \geq (\delta - 2\varepsilon)|X|$. Moreover $|X| \geq M - |\tau|$, and so $d_X(y) \geq (\delta - 2\varepsilon)M - |\tau| \geq (\delta - 3\varepsilon)M$. Since $|X_u| \leq x_u M$, the claim now implies $\sum_u x_u w_u \geq \sum_u \big(d_{X_u}(y)/M - \varepsilon/|\tau|\big) = d_X(y)/M - \varepsilon \geq \delta - 4\varepsilon$.

Let $U(y) = \bigcup_u U_u$. There are at most $\binom{M}{m}^{|\tau|}$ possible choices for this subset of $\bigcup_u V_u$. There are at most $3^{|\tau|}$ possible ways that the edges between y and $U(y)$ can be coloured. Now $|Y| \geq \varepsilon n/2 \geq R(m)3^{|\tau|}\binom{M}{m}^{|\tau|}$. So there is a subset $Z \subset Y$ with $|Z| = R(m)$, such that the sets $U(y)$ are identical for all $y \in Z$ and moreover every $y \in Z$ is joined in the same way to $U(y)$. By the definition of the Ramsey number we can pick a subset of m vertices in Z that span a complete red or a complete blue K_m (in this instance a green edge is thought of as containing both a red and a blue edge). Call this subset U_v. Then U_v together with $\bigcup_{u \in V(\tau)} U_u$ is the vertex set of $m \times \sigma$ for some extension σ of τ, where $\{v\} = V(\sigma) - V(\tau)$. Moreover, by construction, we have $\sum_{u \in V(\tau)} x_u w(uv) = \sum_u x_u w_u \geq \delta - 4\varepsilon$. ■

Lemma 3.13. *Let τ be a type and let $x \in \Delta$ satisfy $\lambda_p(\tau) = x^{\mathrm{t}} W_p(\tau) x$. Let σ be an extension of τ such that $\sum_{u \in V(\tau)} x_u w(uv) \geq \delta > \lambda_p(\tau)$, where $\{v\} = V(\sigma) - V(\tau)$. Then $\lambda_p(\sigma) \geq \delta^2/\big(2\delta - \lambda_p(\tau)\big) > \lambda_p(\tau)$.*

Proof. Given $0 \leq \xi \leq 1$, let $y \in [0,1]^{|\sigma|}$ satisfy $y_u = \xi x_u$ if $u \in V(\tau)$ and $y_v = 1 - \xi$ where $\{v\} = V(\sigma) - V(\tau)$. Then $y \in \Delta\big(|\sigma|\big)$ so $\lambda_p(\sigma) \geq y^{\mathrm{t}} W_p(\sigma) y$. Thus $\lambda_p(\sigma) \geq \xi^2 \lambda_p(\tau) + 2\xi(1 - \xi)\delta$. Taking $\xi = \delta/\big(2\delta - \lambda_p(\tau)\big)$ gives the desired result. ■

A standard lemma relates the minimum degree condition of Lemma 3.11 to the p-weight condition involved in the definition of $\kappa(\mathcal{H})$.

Lemma 3.14. *Let $\mathcal{H}$ be a class of 2-coloured graphs. Then there is a graph sequence (G^n) with $\liminf_{n\to\infty}\left(\delta_p(G^n)/|G^n|\right) = \kappa_p(\mathcal{H})$, such that no G^n contains a member of $\mathcal{H}$.*

Proof. It suffices, for each n, to find a graph G^n with at least n vertices, with minimum degree at least $\left(\kappa_p(\mathcal{H}) - 3/n\right)|G^n|$, and which contains no members of $\mathcal{H}$. This sequence will satisfy $\liminf_{n\to\infty}\left(\delta_p(G^n)/|G^n|\right) \geq \kappa_p(\mathcal{H})$, and it is impossible that the inequality is strict, for then we would have a sequence with $\liminf_{n\to\infty} w_p(G^n)/\binom{n}{2} > \kappa_p(\mathcal{H})$, contradicting the definition of $\kappa_p(\mathcal{H})$.

To find G^n we note that, by the definition of $\kappa = \kappa_p(\mathcal{H})$, there is a some $N > n^2$ and some graph G_N of order N with $w_p(G_N) \geq (\kappa - 1/n)\binom{N}{2}$ and containing no member of $\mathcal{H}$. Construct a maximal sequence of graphs $G_N \supset G_{N-1} \supset \cdots \supset G_M$ such that $|G_j| = j$ and G_{j-1} is obtained from G_j by removing a vertex of degree at most $(\kappa - 2/n)(j-1)$ in G_j. Then $M > n$, for otherwise G_n appears in the sequence, and

$$w_p(G_n) \geq \left(\kappa - \frac{1}{n}\right)\binom{N}{2} - \sum_{j=n+1}^{N}\left(\kappa - \frac{2}{n}\right)(j-1) \geq \frac{1}{n}\binom{N}{2} > \binom{n}{2}.$$

which is impossible. Since the sequence stops at G_M it follows that $\delta_p(G_M) \geq (\kappa - 2/n)(M-1) \geq (\kappa - 3/n)M$, so we may take $G^n = G_M$. ∎

The following approximation theorem is the 2-coloured analogue of [16, Theorem 2] and, given the equivalence of $\kappa_p(\mathcal{H})$ to other parameters as discussed in §1, can be found in [5, Lemma 3.4], [8, Theorem 11] and [12, Theorem 1.1].

Theorem 3.15. *Let $\mathcal{H}$ be a class of 2-coloured graphs. Then*

$$\kappa_p(\mathcal{H}) = \sup\left\{\lambda_p(\tau) \,:\, \textit{no graph in } \mathcal{H} \textit{ is } \tau\textit{-colourable}\right\}.$$

Remark 3.16. This theorem is made redundant by the deeper Theorem 3.25 appearing later, in which the supremum is replaced by a maximum.

Proof. Let $T = \{\tau\text{: no graph in } \mathcal{H} \text{ is } \tau\text{-colourable}\}$. Let $s = \sup\left\{\lambda(\tau) : \tau \in T\right\}$ and $\kappa = \kappa_p(\mathcal{H})$. Observation 3.7 shows that $s \leq \kappa$; the point of the theorem is that equality holds.

Suppose on the contrary that $s < \kappa$. By Lemma 3.14 there is a graph sequence (G^n) with $\liminf_{n\to\infty}\left(\delta(G^n)/|G^n|\right) = \kappa$ such that no G^n contains

any member of $\mathcal{H}$. Ramsey's theorem implies that the sequence (G^n) contains a type τ_1 with a single vertex. By Lemmas 3.11 and 3.13 we can find a sequence of types $\tau_1 \subset \tau_2 \subset \cdots \subset \tau_t$, all contained in (G^n), such that $\lambda_p(\tau_{j+1}) \geq \kappa^2/\big(2\kappa - \lambda_p(\tau_j)\big)$. From this it follows by induction that $\lambda_p(\tau_j) \geq \kappa(1 - 1/j)$. Let $t = \lceil \kappa/(\kappa - s) \rceil + 1$; then $t > \kappa/(\kappa - s)$ and so $\lambda_p(\tau_t) > s$. By the definition of s, some graph $H \in \mathcal{H}$ is τ_t-colourable, which is to say that $H \subset m \times \tau_t$ for some m. But (G^n) contains τ_t and so there is some n for which $m \times \tau_t \subset G^n$. Thus $H \subset G^n$, contradicting the definition of the sequence (G^n) and completing the proof. ■

3.3. Core types

Clearly, if τ is a type, then $\lambda_p(\tau') \leq \lambda_p(\tau)$ for every sub-type τ' of τ (by extending vectors $z' \in \Delta\big(|\tau'|\big)$ by zeros to vectors $z \in \Delta\big(|\tau|\big)$). It might be that there there is some τ' with $\lambda_p(\tau') = \lambda_p(\tau)$. In contexts such as that of Theorem 3.15, where it is only the p-value $\lambda_p(\tau)$ that matters, rather than τ itself, it is evident that we need not consider τ but only τ'. This observation leads naturally to the next definition.

Definition 3.17. Let $0 \leq p \leq 1$ be given. A type τ is called *p-core* if $\lambda_p(\tau') < \lambda_p(\tau)$ for every proper sub-type τ' of τ.

Remark 3.18. Note that a type that is p-core for some p might be p-core for every p but need not be. For example, let $\tau(0,k)$ be the type consisting of k blue vertices joined by green edges (for the more general $\tau(a,b)$, see Definition 5.1). It is easily seen that $\tau(0,k)$ is p-core for all non-zero p; indeed the optimal vector assigns weight $1/k$ to each vertex and $\lambda_p\big(\tau(0,k)\big) = 1 - p/k$. On the other hand, the type of order two consisting of two red vertices joined by a blue edge, is p-core for $p < 1/2$ but not for $p \geq 1/2$.

Second, note that the definition can be stated equivalently in terms of the p-matrix $W_p(\tau)$ of τ, which is the approach of [15, 16, 17]. Indeed, τ is p-core if and only if the matrix $W_p(\tau)$ is *dense* in the sense of those papers.

Lastly, note that this definition *differs* from that in [12] but only in a change of variable (as explained in §2.1 and in [32]); that is, for each p, there is a p', such that τ is p'-core if and only if the corresponding type is p-core in the sense of [12].

On contemplation of the definition, some useful properties of p-core types reveal themselves, and they have been discovered more than once. Consider an optimum vector $x \in \Delta$ such that $x^{\mathrm{t}}Wx = \lambda_p(\tau)$, where $W = W_p(\tau)$. In a graph in $\mathcal{Q}(\tau)$ of maximum p-weight, the class sizes are proportional to the entries of x (*q.v.* Remark 3.6). But the degrees in a graph of maximum p-weight are evidently more or less equal (if not, move a vertex from one class to another, as done in [12, Lemma 5.3], [15, §2]). This means $Wx = \lambda_p(\tau)e$, where e is the all one vector. We make note of this fact.

Fact 3.19. *Let τ be a p-core type with p-weight $\lambda = \lambda_p(\tau)$. Let $x \in \Delta$ satisfy $x^{\mathrm{t}}Wx = \lambda$, where $W = W_p(\tau)$. Then $Wx = \lambda e$. In other words, $x_u w_p(u) + \sum_{uv \in E(\tau)} x_v w_p(uv) = \lambda$ holds for all $u \in V(\tau)$.*

The definition of p-core means that no co-ordinate of x can be zero. Thus by shifting weight from one vertex to another it can be seen that τ has no twin vertices ([12, Lemma 5.2], [15, §2]). Indeed, if $w_p(uv) \leq w_p(u) = w_p(v)$, define the vector y by $y_u = -x_u$, $y_v = x_u$, $y_i = 0$ otherwise; then $y^{\mathrm{t}}e = 0$ and $Wy = 0$. Let $z = x + y$. Then $z \in \Delta$ and $z^{\mathrm{t}}Wz = x^{\mathrm{t}}Wx + 2y^{\mathrm{t}}Wx + y^{\mathrm{t}}Wy = \lambda$, giving an optimal vector z with a zero co-ordinate, a contradiction. We note this fact too.

Fact 3.20. *Let τ be a p-core type and let u and v be two vertices with $w_p(u) = w_p(v)$. Then $w_p(uv) > w_p(u)$.*

The converse of Fact 3.19 clearly holds also, since if τ is p-core and $Wx = \lambda e$ then x is optimal. So convex linear combinations of optimal vectors are themselves optimal, and hence there must be a unique optimum (else we could create one with a zero entry). Then W is invertible (else we could perturb the optimum optimally), so $x = \lambda W^{-1}e$ and $\lambda = 1/e^{\mathrm{t}}W^{-1}e$ ([17, Lemma 2], [8, Lemma 10]). However, we make no use of these properties of p-core types. Neither shall we use Sidorenko's elegant characterization of their matrices [36], but it bears repeating all the same.

Proposition 3.21 (Sidorenko [36]). *A type τ is p-core if and only if its p-matrix $W = W_p(\tau)$ satisfies*

- *W is non-singular and all components of $W^{-1}e$ are positive, and*
- *$y^{\mathrm{t}}Wy < 0$ for every non-zero vector y with $e^{\mathrm{t}}y = 0$.*

The next lemma will be a very useful tool in handling p-core types, despite its rather clumsy statement.

Lemma 3.22. *Let τ be a p-core type where $p \leq 1/2$ and let $x \in \Delta(|\tau|)$ satisfy $x^{\mathrm{t}} W_p(\tau) x = \lambda_p(\tau)$. Let B be the set of blue vertices of τ, and suppose that all the edges of τ incident with B are green. Let σ be an extension of τ such that $\sum_{u \in V(\tau)} x_u w(uv) > \lambda_p(\tau)$, where $\{v\} = V(\sigma) - V(\tau)$.*

Suppose now that τ' is a p-core sub-type of σ with $\lambda_p(\tau') = \lambda_p(\sigma)$. Then $B \subset V(\tau')$, and all the edges of τ' incident with B are green.

Proof. Let $b \in B$. Fact 3.19 implies that $x_b w_p(b) + \sum_{bu \in E(\tau)} x_u w(bu) = \lambda_p(\tau)$. But $w(bu) = 1$ for all $bu \in E(\tau)$. So $\sum_{u \in V(\tau)} x_u w(uv) > \lambda_p(\tau)$ is possible only if $w(bv) > w(b)$, that is, the weight of the edge bv is greater than that of the vertex b itself. But b is blue and $q \geq 1/2$, so bv must be green. Thus b is joined by green to every other vertex in σ. Hence every vertex of σ incident with B is green.

It remains to show that $B \subset V(\tau')$. Now by Lemma 3.13, $\lambda_p(\tau') = \lambda_p(\sigma) > \lambda_p(\tau)$. Since τ is p-core this means τ' cannot be contained in τ and so $v \in V(\tau')$. Suppose that some vertex $b \in B$ is not contained in τ'. Clearly the type $\tau' - v + b$, which is a sub-type of τ, must have p-value at least that of τ', because b is blue and is joined to every other vertex by green. But this would again contradict τ being p-core: thus $b \in V(\tau')$. ∎

The next theorem, which is of practical significance, is the main result showing that the edge colours in a p-core type are very restricted; they must be green, unless they join two vertices of the same colour and this colour has the smaller weight. The proof is based on that of Theorem 5 in [17] (announced in [15, Theorem 3]), which is the analogous theorem for undirected multigraphs with edge multiplicity $m \leq 2$. A suitable analogue for higher multiplicites is unknown, so this is one of the properties of multiplicity $m \leq 2$ multigraphs that holds for 2-coloured multigraphs, as discussed in §2.3.

Theorem 3.23. *Let τ be a p-core type. Then all edges of τ are green, apart from*

- *if $p < 1/2$, when some edges joining two red vertices might be blue, or*
- *if $p > 1/2$, when some edges joining two blue vertices might be red.*

Remark 3.24. The fact that all edges must be green when $p = 1/2$ accounts for the simplicity of the study of random hereditary properties with edge probability $1/2$ in [11] as compared with the general case in [12].

Proof. By Fact 3.20 and by symmetry, it is enough to assume that $p < 1/2 < q$ and to show that all edges meeting a blue vertex are green. Let B be the set of blue vertices of τ. Certainly the edges within B are green, by Fact 3.20. Let ρ_0 be the sub-type of τ spanned by B. Then $\rho_0 = \tau\big(0, |B|\big)$ (see Remark 3.18) so ρ_0 is p-core.

We shall show that, given a p-core proper sub-type $\rho \subset \tau$ which contains B and in which all edges incident with B are green, then there is another p-core sub-type $\rho' \subset \tau$, also containing B and in which all edges meeting B are green, and such that $\lambda_p(\rho) < \lambda_p(\rho')$. Applying this process repeatedly, starting with ρ_0, we generate a sequence of sub-types of strictly increasing p-value, which must therefore reach τ after a finite number of steps, completing the proof.

Let ρ be a p-core sub-type as described and let $x \in \Delta\big(|\rho|\big)$ be an optimum vector. Let (G^n) be a sequence of graphs in $\mathcal{Q}(\tau)$ of maximal weight; then $\liminf \big(\delta_p(G^n)/|G^n|\big) = \lambda_p(\tau)$ (see Remark 3.6). Since τ is p-core, and no co-ordinate of an optimum vector is zero, the construction of G^n shows that the sequence (G^n) contains τ, and hence it certainly contains ρ.

If $\rho \neq \tau$, then Lemma 3.11 applied to ρ shows that (G^n) contains an extension σ of ρ with $\sum_{u \in V(\rho)} x_u w(uv) \geq \lambda_p(\tau) > \lambda_p(\rho)$, where v is the vertex added to ρ to make σ. Let ρ' be a p-core sub-type of σ with $\lambda_p(\rho') = \lambda_p(\sigma)$. Then Lemma 3.22, applied to ρ and ρ', shows that $B \subset V(\rho')$ and that all edges of ρ' incident with B are green.

All that remains is to show that ρ' is actually a sub-type of τ. Now σ is contained in (G^n) and so ρ' is also. In particular, $\big(|\tau| + 1\big) \times \rho' \subset G^n$ for some n. Denote the classes of $\big(|\tau| + 1\big) \times \rho'$ by U_u, $u \in V(\rho')$. By the construction of (G^n) we may pick, for each u, a vertex $f(u) \in V(\tau)$ such that two vertices of U_u lie in the same class $V_{f(u)}$ of G^n. But then u must have the same colour as $f(u)$. If now $f(u') = f(u)$ for some other vertex $u' \in V(\rho')$, then the edge uu' is the same colour as both u and u', contradicting Fact 3.20. So the map $f : V(\rho') \to V(\tau)$ must be injective, and this means $\rho' \subset \tau$. ∎

Theorem 3.23 comes in very handy when calculating $\kappa_p(H)$ for 2-coloured multigraphs H, as we shall see when we look at examples in §5.

3.4. Exactness

We are now in a position to prove the principal positive result of §3. This corresponds to the main theorem of Brown, Erdős and Simonovits [15] for multigraphs with multiplicty $m \leq 2$, on whose proof the proof here is based.

Theorem 3.25. *Let $\mathcal{H}$ be a class of 2-coloured multigraphs and let $0 \leq p \leq 1$. Then there is a p-core type τ such that no $H \in \mathcal{H}$ is τ-colourable, and $\lambda_p(\tau) = \kappa_p(\mathcal{H})$. That is,*

$$\kappa_p(\mathcal{H}) = \max\left\{ \lambda_p(\tau) \,:\, \textit{no graph in } \mathcal{H} \textit{ is } \tau\textit{-colourable} \right\}.$$

Remark 3.26. It is possible that infinitely many types satisfy the conditions of the theorem, as will be shown in §3.5.

Proof. By symmetry, we may assume $p \leq 1/2$. Select a graph sequence (G^n) as given by Lemma 3.14, with $\liminf_{n\to\infty} \left(\delta(G^n)/|G^n|\right) = \kappa_p(\mathcal{H})$, such that no G^n contains a member of $\mathcal{H}$. If σ is a p-core type contained in (G^n) then, by Theorem 3.23, the blue vertices of σ are joined to each other by green edges. So if there are k of these vertices then they span the type $\tau(0,k)$ (as mentioned in Remark 3.18). If $H \in \mathcal{H}$ then H is $\tau\big(0,|H|\big)$-colourable so $k < |H|$. Hence the number $t = \max\left\{ k \,:\, \tau(0,k) \text{ is contained in } (G^n) \right\}$ exists and is finite.

The theorem will be proved if we show that there is some τ contained in (G^n) with $\lambda_p(\tau) = \kappa_p(\mathcal{H})$. We shall assume, then, that no such τ exists, and derive that $\tau(0,t+1)$ is contained in (G^n), a contradiction that completes the proof.

Let τ be a p-core type contained in (G^n) such that $\tau(0,t) \subset \tau$. Choose $x \in \Delta$ so that $x^{\mathrm{t}} W_p(\tau) x = \lambda_p(\tau)$, and apply Lemma 3.11 to obtain an extension σ of τ that is contained in (G^n). Let τ' be a p-core sub-type of σ with $\lambda_p(\tau') = \lambda_p(\sigma)$. By Lemma 3.13, $\lambda_p(\tau') > \lambda_p(\tau)$, and by Lemma 3.22, $\tau(0,t) \subset \tau'$. Since we never come across a τ with $\lambda_p(\tau) = \kappa_p(\mathcal{H})$, this means we can construct an infinite sequence $\tau(0,t) = \tau_0, \tau_1, \tau_2, \ldots$ of p-core types, each contained in (G^n), each containing $\tau(0,t)$, and of strictly increasing p-value; in particular they are distinct and so their orders are unbounded.

Let m be a positive integer. Choose some τ_j in the sequence with $|\tau_j| \geq t+m$ and let ρ be a sub-type of τ_j containing $\tau(0,t)$ plus m other vertices. By Theorem 3.23 applied to τ_j, all the edges of ρ meeting $\tau(0,t)$ are green, and the other edges are blue or green. Since ρ is contained in (G^n), there is some n with $m \times \rho \subset G^n$. Now the t classes of $m \times \rho$ corresponding

to the sub-type $\tau(0,t) \subset \rho$ give a copy of $m \times \tau(0,t)$; by picking m other vertices, one in each of the m remaining classes of $m \times \rho$, we obtain m further vertices joined to each other by blue or green edges and joined to $m \times \tau(0,t)$ by green edges. This construction shows that $m \times \tau(0,t+1) \subset m \times \rho$. and so $m \times \tau(0,t+1) \subset G^n$.

Therefore $\tau(0,t+1)$ is contained in (G^n), and we are done. ■

3.5. Stability and Finiteness

Unlike the analogous theorem for multigraphs with multiplicity $m \leq 2$, there can be an infinite number of types that satisfy the conditions of Theorem 3.25. To show this, we adapt an example of Sidorenko [36], constructed to show the corresponding fact for multigraphs with $m \geq 3$.

Theorem 3.27. *Let $l \geq 4$ be an integer and let $p = 1/l$. Let H be the complete two-coloured multigraph of order $l+1$ consisting of a red star $K_{1,l}$ with all other edges being blue. Then there are infinitely many p-core types τ such that H is not τ-colourable and $\lambda_p(\tau) = \kappa_p(H)$.*

Proof. It will be shown later in Example 5.5 that $\kappa_p(H) = 1 - 1/l$. If τ is any type whose vertices are red and in which the green edges form an $(l-2)$-regular graph, the other edges being blue, then H is not τ-colourable. Theorem 5.6 shows that any such τ for which the green edges form a connected graph is p-core when $p = 1/l$, and $\lambda_p(\tau) = 1 - 1/l = \kappa_p(H)$. ■

Thus there is no hope of proving a stability result along the lines that an extremal sequence for H must contain graphs whose structure is close to one of a finite number of possible structures. We mention, though, that Alon and Stav [6] have proved certain stability results related to $\kappa_{1/2}(\mathcal{H})$.

Another finiteness property that holds for multigraphs with multiplicity $m \leq 2$ is that every family contains a finite subfamily with the same extremal density [17, Theorem 3]. This too fails for $m \geq 4$, as Rödl and Sidorenko [35] showed.

The example of Rödl and Sidorenko is based on the usual *Lagrangian* function of a multigraph. Given a multigraph G and a vector $x \in \Delta\big(|G|\big)$ we define the quantity $\lambda(G,x) = \sum_{uv \in E(G)} x_u x_v$. The Lagrangian of G is then $\lambda(G) = \max\big\{\lambda(G,x) : x \in \Delta\big\}$. The central element of their

argument is a (wonderfully elegant) proof that, for each integer $d \geq 3$ and each integer k, there exists a multigraph F, in which each pair of vertices is joined by either d or $d+1$ edges, such that $\lambda(F) > d$, but for each sub-multigraph $F' \subset F$ with $|F'| \leq k$, $\lambda(F') < d$ holds. (This multigraph is the one in Theorem 2.4 of [35], its properties being given in Corollary 2.3 and the proof of Theorem 2.7.)

Theorem 3.28. *Let $l \geq 5$ be an integer and let $p = 1/l$. Then there exists a family $\mathcal{H}$ of 2-coloured multigraphs for which there is no finite sub-family $\mathcal{H}_0 \subset \mathcal{H}$ with $\kappa_p(\mathcal{H}_0) = \kappa_p(\mathcal{H})$.*

Proof. For each complete 2-coloured multigraph H we define the (ordinary, uncoloured) multigraph $H_{\rm m}$ to be the multigraph with vertex set $V(H)$, such that two vertices $u, v \in V(H)$ are joined in $H_{\rm m}$ by 0, $l-2$ or $l-1$ edges according as the edge uv is red, blue or green. It is important to the argument that if $H \subset H'$ then $H_{\rm m} \subset H'_{\rm m}$ (though the converse does not hold).

Observe that if H is a complete 2-coloured multigraph and $x \in \Delta\big(|H|\big)$ is the vector with all co-ordinates equal to $1/|H|$, then $\lambda(H_{\rm m}, x)$ equals $2(l-2)/|H|^2$ times the number of blue edges in H plus $2(l-1)/|H|^2$ times the number of green edges. Since $p = 1/l$ we have $w_p(H) = p\binom{|H|}{2} + \lambda(H_{\rm m}, x)\frac{|H|^2}{2l}$.

Let $\mathcal{H} = \big\{H : \lambda(H_{\rm m}) > l-2\big\}$. By the definition of $\kappa_p(\mathcal{H})$ there is a sequence (G^n) of complete 2-coloured multigraphs, none of which contains any member of $\mathcal{H}$, with $|G^n| = n$ and with $w_p(G^n) = \big(\kappa_p(\mathcal{H}) + o(1)\big)\binom{n}{2}$. But $w_p(G^n) \leq \big(p + \lambda(G^n_{\rm m})/l + o(1)\big)\binom{n}{2}$. Since, certainly, $G^n \notin \mathcal{H}$, we have $\lambda(G^n_{\rm m}) \leq l-2$. Thus $\kappa_p(\mathcal{H}) \leq (l-1)/l = q$.

Suppose now that $\mathcal{H}_0 \subset \mathcal{H}$ and $\mathcal{H}_0$ is finite. Let $k = \max\big\{|H| : H \in \mathcal{H}_0\big\}$. Let F be the multigraph of Rödl and Sidorenko defined above, taking $d = l-2$. Then $F = G_{\rm m}$ for some complete 2-coloured multigraph G, none of whose edges are red. Let τ be the type obtained by colouring each vertex of G red.

We claim that no $H \in \mathcal{H}_0$ is τ-colourable. For suppose instead that $H \subset m \times \tau$. Since $|H| \leq k$, there is a sub-type $\sigma \subset \tau$, with $|\sigma| \leq k$, such that $H \subset m \times \sigma$. As noted at the beginning of the proof, this means $H_{\rm m} \subset (m \times \sigma)_{\rm m}$, which in turn implies $\lambda(H_{\rm m}) \leq \lambda\big((m \times \sigma)_{\rm m}\big)$. The vertices of σ span a sub-multigraph $F' \subset F$ with $|F'| \leq k$, so $\lambda(F') < l-2$ by the property of F. But $(m \times \sigma)_{\rm m}$ is obtained from F' by replacing each vertex of F' by a class of m vertices, with vertices in different classes

being joined to each other by edges of multiplicity equal to the multiplicity between the original vertices of F'. It is a standard and simple property of the Lagrangian that $\lambda\big((m \times \sigma)_{\mathrm{m}}\big) = \lambda(F')$. So $\lambda(H_{\mathrm{m}}) < l-2$, contradicting $H \in \mathcal{H}_0$.

Because no $H \in \mathcal{H}_0$ is τ-colourable, Observation 3.7 implies $\kappa_p(\mathcal{H}_0) \geq \lambda_p(\tau)$. Let $x \in \Delta\big(|\tau|\big)$. Since $p = 1/l$, and by the construction of $G_{\mathrm{m}} = F$, we have $x^{\mathrm{t}} W_p(\tau) x = p\big(1+\lambda(F,x)\big)$. Thus $\lambda_p(\tau) = p\big(1+\lambda(F)\big) > p(l-1) = q$. Hence $\kappa_p(\mathcal{H}_0) > \kappa_p(\mathcal{H})$, completing the proof. ■

Remark 3.29. It is interesting to note that Theorem 3.28 fails for $p = 1/2$; in this case, there is always a suitable finite $\mathcal{H}_0 \subset \mathcal{H}$. This is because the only p-core types with $p = 1/2$ are the types $\tau(a,b)$, comprising a red and b blue vertices joined to each other by green edges (see Example 5.2), and, for any class $\mathcal{H}$, $\kappa_{1/2}(\mathcal{H}) = 1 - 1/2t$ where t is the maximum value of $a+b$ for which there is a $\tau(a,b)$, such that no $H \in \mathcal{H}$ is $\tau(a,b)$-colourable. We can, for each j with $0 \leq j \leq t+1$, choose some $H_j \in \mathcal{H}$ which is $\tau(t+1-j, j)$-colourable. Then, putting $\mathcal{H}_0 = \{H_0, \ldots, H_{t+1}\}$ we obtain a finite sub-family $\mathcal{H}_0 \subset \mathcal{H}$ with $\kappa_{1/2}(\mathcal{H}_0) = \kappa_{1/2}(\mathcal{H})$.

3.6. Incomplete extremal graphs

As promised at the start of §1, we now look briefly at what can be said about the extremal function for 2-coloured multigraphs if we do not constrain the extremal graphs to be complete. That is, we study the function

$$\mathrm{ex}_p(H,n) = \max\big\{\, w_p(G) \,:\, |G| = n,\ H \not\subset G\big\}$$

rather than $\mathrm{kex}_p(H,n)$, and we define $\mu_p(H) = \lim_{n\to\infty} \mathrm{ex}_p(H.n)\binom{n}{2}^{-1}$.

In many small cases where $\kappa_p(H)$ and $\mu_p(H)$ can be evaluated, they are equal, though this is not always the case (see Example 5.10) and there is no immediate reason to think that equality is common. However if, say, H has only blue and green edges, then $\kappa_p(H) = \mu_p(H)$, because the addition of red edges to an extremal graph which is not complete will not create a copy of H.

To extend the methods of §3 to accommodate incomplete extremal graphs, we need to extend the notion of a type to allow white vertices and edges, which have weight zero. If τ is such an extended type then the

definition of $m \times \tau$ would preclude any edges within classes corresponding to white vertices or between two classes corresponding to a white edge.

The results of sections §3.1 and §3.2 would then still hold good in this extended context, because of their generic nature; we define $\lambda_p(\tau)$ for an extended type in the same way as for an ordinary type, though of course the p-matrix $W_p(\tau)$ might contain some zero entries. For the same reason the initial simple results in §3.3 would still hold, such as Facts 3.19 and 3.20, and of course Proposition 3.21.

After that, things start to go wrong. Lemma 3.22 still holds good as stated, but the proof does not extend to showing that the edges at each vertex of a p-core type have weight greater than that of the vertex. So Theorem 3.23 is weakened to stating that (if $p > 1/2$) all edges meeting a blue vertex are green, together with whatever information is supplied by Fact 3.20. The main casualty is, of course, Theorem 3.25; the possibility of non-blue edges between non-blue vertices destroys the heart of the proof.

The counterexamples of §3.5 still stand. Theorem 5.6 still gives infinitely many graphs with the same value of λ_p and, because $\mu_p(H) = \kappa_p(H)$ for the graph H of Theorem 3.27 (see [31]), the theorem remains true. Theorem 3.28 goes through essentially without change: the graphs H_{m} are defined in the same way except that white edges, like red edges, contribute nothing to H_{m}, and the equation given for $w_p(H)$ remains true as an upper bound, which is all that is needed.

An application of the parameter μ_p is discussed in the next section.

4. Graph Games and Incomplete Extremal Graphs

There is a rich and fascinating connection between games on graphs and Ramsey theory, described in detail by Beck [9]. One such game, played on the edges of a complete graph K_n by two players, was studied by Richer [34]. Before the game, some 2-coloured complete graph H is specified, which has no green edges. The first player, when it is his turn, selects some fixed number of edges of K_n and colours them red or blue. The second player, in turn, selects an edge and labels it $+$. The first player wins if, at the end of the game (when all of K_n is coloured), there is a copy of H in K_n all of whose edges are labelled $+$. Whether the second player can win depends on how many edges the first player must colour at each turn.

A simpler version of this game is one where the first player selects outright a red/blue colouring of K_n and the second player then labels some prescribed number of the edges. The first player wins if the second player cannot avoid a copy of H within the labelled edges. In this form of the game, the parameter of interest is $\ell(H)$, which is the minimum value of ℓ such that the first player can win if n is large and the second player must label $\ell\binom{n}{2}$ edges.

More generally, let $\ell_p(H,n)$ be the minimum number of edges the second player must be made to label in order that the first player can find a winning colouring with $\lceil p\binom{n}{2}\rceil$ red edges, and let $\ell_p(H) = \lim_{n\to\infty} \ell_p(H,n)\binom{n}{2}^{-1}$ (the existence of this limit is shown by the proof of the next theorem). Then we can define $\ell(H) = \min_p \ell_p(H)$.

It turns out that $\ell(H)$ behaves similarly to the parameter ed $\big(\mathrm{Forb}\,(H)\big)$ of §2.2 except that, rather than $\kappa_p(H)$, it is the "incomplete" extremal function $\mu_p(H)$ that is involved (see §3.6).

Theorem 4.1. *For the game above,* $\ell_p(H) = \mu_p(H)$ *and so* $\ell(H) = \min_p \mu_p(H)$.

Proof. Rather than dot every 'i' and cross every 't' we shall sketch the proof in the spirit of §2.2; however, unlike in the case of edit distance, the application of Szemerédi's Lemma is standard, and the reader familiar with this Lemma and with standard facts about random graphs, such as are frequently used in the works on hereditary property and edit distance cited in this paper, can easily fill in the details if minded to do so. We suppose that n is large, that the first player must colour $p\binom{n}{2}$ edges red, and that the second player must label $\ell\binom{n}{2}$ edges.

If $\ell < \mu_p(H)$, the second player chooses a 2-coloured graph F of order n containing no H and with $w_p(F) \geq \ell\binom{n}{2}$. Remember that the underlying graph of F need not be complete, but the missing edges just have weight zero. Now, whatever 2-coloured K_n is chosen by the first player, the second player finds a bijection of $V(F)$ to $V(K_n)$ which maximizes the number of edges of K_n agreeing in colour with the corresponding edge of F; a green edge of F will agree with an edge of K_n of either colour, whereas a white (absent) edge of F agrees with neither colour. Since K_n has $p\binom{n}{2}$ red edges, any given red edge of F has probability p of agreeing with its corresponding edge in K_n, and so the expected number of agreeing edges in a random bijection is $w_p(F)$. Hence by labelling the agreeing edges in a bijection of maximum agreement, the second player labels at least $\ell\binom{n}{2}$ edges, and

these all lie in F so contain no copy of H. Hence the second player wins if $\ell < \mu_p(H)$, and so $\ell_p(H,n) \geq \mu_p(H)\binom{n}{2}$.

On the other hand, if $\ell > \mu_p(H) + 6\varepsilon$, then the first player selects a random colouring of K_n with red probability p. The second player then labels $\ell\binom{n}{2}$ edges; let L_r and L_b be the red and blue labelled subgraphs. Find a partition of $V(K_n)$ that is a Szemerédi partition simultaneously for both L_r and L_b, with $k > 2/\varepsilon$ classes and $\eta = \eta(H,\varepsilon)$-regular pairs (this exists, see [29, p. 306]). Form the 2-coloured reduced graph R of order k, whose red edges represent regular pairs of density $> \eta$ in L_r and whose blue edges behave similarly with respect to L_b. The number of labelled edges is at most the edges within classes plus the edges in irregular pairs plus the edges in low density pairs plus the edges in regular pairs. Since the colouring is random, almost surely there are at most $(p+\varepsilon)(n/k)^2$ red edges between any pair of n/k sets in the colouring, and so in particular at most this many edges of L_r in a pair corresponding to a red edge of R. Hence the number of labelled edges is at most $4\varepsilon\binom{n}{2} + w_p(R)(n/k)^2$. Thus $w_p(R) \geq (\mu_p(H)+\varepsilon)\binom{k}{2} \geq \mathrm{ex}(H,k)$ if k was chosen large enough, and so R contains a copy of H. By the standard property of Szemerédi partitions, H can now be constructed in $L_r \cup L_b$, and the first player wins. Hence $\ell_p(H,n) \leq (\mu_p(H)+6\varepsilon)\binom{n}{2}$ for large n. ∎

The proof shows that random colourings are always a good choice for the first player, though there are sometimes other choices possible. The fact that random colourings are good is made even clearer in the more general game studied by Richer, for which some number $0 \leq t \leq \binom{|H|}{2}$ is also specified before the game, and the first player wins if the second player is forced to label at least t of the edges of some copy of H. (So up till now we have been playing this game with $t = \binom{|H|}{2}$.) Let $\mathcal{H}$ be the class of graphs on $V(H)$ comprising t edges of H. Then the first player wins if the second player must label at least $\min_p \mu_p(\mathcal{H})$ edges and otherwise the second player wins. The above proof works for the general game without change until the very end, where we need to observe that R contains some graph H' comprising t edges of H. As before, the pairs corresponding to edges of H' are regular in L_r or L_b as needed. But the other pairs correponding to edges of $H \setminus H'$ are regular in the red or blue edges of K_n because the colouring is random. So we can construct a copy of H in K_n in which t edges, the edges in H', are labelled.

In [34] Richer gives many examples and much information about $\ell(H)$ and $\ell(\mathcal{H})$ for the corresponding families $\mathcal{H}$, and in particular all values for

$|H| \leq 4$. Exact values of $\mathrm{ex}\,(H, n)$ are given for all n and all H with $|H| \leq 3$ and no isolated vertices; there are ten such graphs. His work necessarily contains values for $\mu_p(H)$ in many instances. For example, let H be the 2-coloured complete graph of order 4 having one green edge, a disjoint red edge and a blue $K_{2,2}$ in between. Then $\kappa_p(H) = \mu_p(H) = \max\{1 - q/2, 1 - pq\}$, the extreme types being either two red edges joined by a green edge, or one blue and one red vertex joined by a green edge.

The most noteworthy question left open by Richer is the following conjecture, having the flavour of the Erdős–Stone theorem.

Conjecture 4.2 (Richer [34]). Let H be a 2-coloured multigraph with no green edges. Then there exists some value of p such that

$$\kappa_p(H) \leq \mu_p(H) \leq 1 - \frac{1}{|H| - 1} \qquad \text{if } |H| \text{ is odd.}$$

Notice that the conjecture is best possible, insofar as if H is complete then a complete $\big(|H| - 1\big)$-partite graph coloured green contains no copy of H. This colouring corresponds to the extended type with $|H| - 1$ white vertices joined to each other by green edges.

The conjecture implies $\kappa_p(H) \leq \mu_p(H) \leq 1 - 1/|H|$ for $|H|$ even. The stronger inequality $\kappa_p(H) \leq \mu_p(H) \leq 1 - 1/\big(|H| - 1\big)$ does not always hold when $|H|$ is even; for example, if $H_r = K_{t,t}$ and $H_b = \overline{H}_r = 2K_t$ then $\min_p \kappa_p(H) = 3/4$ when $t = 2$ and $\min_p \kappa_p(H) > 1 - 1/(2t - 1)$ when $t = 3$ and $t = 4$ (see §5.3). However the stronger inequality might hold when $|H|$ is even and large, perhaps if $|H| > 8$. There is a cognate conjecture of Diwan and Mubayi [20, Conjecture 3] which is interesting in this respect. It asserts that, if H is a 2-coloured multigraph with no green edges, then a 2-coloured multigraph G on n vertices with $\min\big\{e(G_r), e(G_b)\big\} > \big(1 - 1/\big(|H| - 1\big)\big)n^2/2$ will contain H unless $|H| = 2t$ with $t = 2$, 3 or 4, H is complete, and the edges of one colour form a spanning subgraph whose components are all balanced complete bipartite graphs.

Richer proved his conjecture for $|H| \leq 5$, and established the general bound $\min_p \mu_p(H) \leq 1 - 3/\big(5|H| - 5\big)$. Further results are given in [31].

5. Examples

In this section we discuss the value of $\kappa_p(H)$ for various graphs H. To find $\kappa_p(H)$ we find a p-core type τ whose existence is guaranteed by Theorem 3.25, such that H is not τ-colourable and $\lambda_p(\tau) = \kappa_p(H)$.

We divide the examples into two classes, those in which the edges of τ are all green, which we call the simple examples, and those in which red or blue edges can appear. Theorem 3.23 helps to reduce the number of types that need be considered, but even with this theorem there can remain significant difficulties.

We begin by giving names to those important core types which contain only green edges.

Definition 5.1. The type $\tau(a, b)$ has a red vertices and b blue vertices, joined by green edges.

Thus $\mathcal{Q}\big(\tau(a, b)\big)$ consists of all complete 2-coloured multigraphs whose vertices can be split into $a + b$ classes, a of which span only red edges and the other b spanning only blue edges. These are the 2-coloured analogues of the hereditary properties $\mathcal{P}(a + b, a)$ found in [11].

The vector $x \in \Delta$ which assigns vertex weights in the proportions $p :: q$ to red and blue vertices is optimal by Fact 3.19. It follows that $\lambda_p\big(\tau(a, b)\big) = 1 - pq/(ap + bq)$, and that τ is p-core if $p \neq 0, 1$.

We shall generally assume that $p \neq 0, 1$ since these extreme cases are covered in §3.1. On occasion we might also assume that $p \neq 1/2$, since $\kappa_{1/2}(\mathcal{H})$ is treated in Example 5.2 for all $\mathcal{H}$.

5.1. Simple examples

Example 5.2. Suppose $p = 1/2$. Then, by Theorem 3.23, p-core types contain only green edges, so the types $\tau(a, b)$ are the only such types. Now $\lambda_{1/2}\big(\tau(a, b)\big) = 1 - 1/2(a+b)$. Therefore, for any class $\mathcal{H}$, $\kappa_{1/2}(\mathcal{H}) = 1 - 1/2t$ where t is the maximum value of $a + b$ for which there is a $\tau(a, b)$, such that no $H \in \mathcal{H}$ is $\tau(a, b)$-colourable. (This value of $\kappa_{1/2}(\mathcal{H})$ appears also in [5]. The number t is the colouring number mentioned after Definition 3.3; $t+1$ is the binary chromatic number studied in [7], where this result is also given.)

Example 5.3. Let τ be a p-core type and let $\mathcal{H}(\tau)$ be all graphs that are not τ-colourable; that is, $\mathcal{H}(\tau)$ is exactly those graphs not in $\mathcal{Q}(\tau)$. So $\lambda_p(\tau) \le \kappa_p\big(\mathcal{H}(\tau)\big)$ by Observation 3.7. On the other hand, any sequence of extremal graphs for the class $\mathcal{H}(\tau)$ must, by definition, lie in $\mathcal{Q}(\tau)$ and so $\kappa_p\big(\mathcal{H}(\tau)\big) \le \lambda_p(\tau)$ by Remark 3.6 and, indeed, $\mu_p\big(\mathcal{H}(\tau)\big) \le \lambda_p(\tau)$ for the same reason. Thus $\kappa_p\big(\mathcal{H}(\tau)\big) = \mu_p\big(\mathcal{H}(\tau)\big) = \lambda_p(\tau)$ for all p.

In particular, $\kappa_p\big(\mathcal{H}\big(\tau(a,b)\big)\big) = \mu_p\big(\mathcal{H}\big(\tau(a,b)\big)\big) = 1 - pq/(ap+bq)$. Note that $\kappa_p\big(\mathcal{H}\big(\tau(a,b)\big)\big)$ is minimized when $p = \sqrt{b}/(\sqrt{a}+\sqrt{b})$, and then $\kappa_p\big(\mathcal{H}\big(\tau(a,b)\big)\big) = 1 - 1/\big(\sqrt{a}+\sqrt{b}\big)^2$ or, equivalently, $\operatorname{ed}\big(\mathcal{P}(a,b)\big) = 1/\big(\sqrt{a}+\sqrt{b}\big)^2$, as proved in [5].

The following bound will prove useful for some more elementary examples.

Lemma 5.4. *Let τ be a p-core type whose vertices are all red, each joined to at most d others by green edges. Then $\lambda_p(\tau) \le \max\big\{q, 1-q/(d+1)\big\}$.*

Proof. Let the maximum number of green edges meeting any vertex be $m \le d$. Then $|\tau| \ge m+1$. Let $x \in \Delta\big(|\tau|\big)$ be a vector with $\lambda_p(\tau) = x^{\mathrm{t}} W_p(\tau) x$. Let G be the graph with vertex set $V(\tau)$ whose edges are the green edges of τ and let $d(u)$ denote the degree of the vertex u in G. By Fact 3.19 we have $\lambda_p(\tau) = px_u + (1-x_u)q + p\sum_{uv\in E(G)} x_v$ for every $u \in V(\tau)$. Summing this equation over all u we obtain $|\tau|\lambda_p(\tau) = p - q + q|\tau| + p\sum_u d(u)x_u \le p - q + q|\tau| + pm$. Hence $\lambda_p(\tau) \le q + \big(p(m+1) - q\big)/|\tau|$. If $p(m+1) - q \le 0$ we have $\lambda_p(\tau) \le q$; if $p(m+1) - q > 0$, we have $\lambda_p(\tau) \le q + \big(p(m+1)-q\big)/(m+1) = 1 - q/(m+1) \le 1 - q/(d+1)$. ∎

Example 5.5. Let $H_r = K_a + \overline{K}_b$ and $H_b = \overline{H}_r = K_b$; that is, $|H| = a+b$, H is complete with no green edges, and consists of a red K_a joined to a blue K_b by red edges. In particular, if $a = 1$, the red graph H_r is a star $K_{1,|H|-1}$. Note that H contains a blue K_b and a red K_{a+1}, so is neither $\tau(b-1,0)$-colourable nor $\tau(0,a)$-colourable. Hence $\kappa_p(H) \ge \max\big\{1 - p/a, 1 - q/(b-1)\big\}$.

Let τ be such that H is not τ-colourable and $\lambda_p(\tau) = \kappa_p(H)$, as given by Theorem 3.25. Then τ cannot have both a red vertex and a blue vertex for, if so, they would be joined by a green edge by Theorem 3.23, and H would be τ-colourable. So the vertices of τ are all the same colour. If they are all red, the edges of τ are green and blue, so no vertex can be joined to $b-1$ others by green (else H is τ-colourable); by Lemma 5.4, this means

$\lambda_p(\tau) \leq 1 - q/(b-1)$. If the vertices of τ are all blue, they are joined to each other by red and green, and none can be joined to a others by green (else H is τ-colourable), so $\lambda_p(\tau) \leq 1-p/a$ by the complementary version of Lemma 5.4. Thus we have shown that $\kappa_p(H) = \max\{1-p/a, 1-q/(b-1)\}$.

The edit distance of H with $(a,b) = (1,3)$ was computed in [5] as an example where the minimum value of p is not $1/2$ (in this case it is $1/3$). The edit distance of H for all (a,b) was given in [8].

A more precise estimate of $\mathrm{kex}(H,n)$ is given in [31].

As discussed in §3.5, the graphs of Example 5.5 can sometimes have infinitely many τ satisfying Theorem 3.25. For example, if $(a,b) = (1,4)$, that is, H_r is the star $K_{1,4}$, then $\kappa_{1/4}(H) = 3/4$, and the next theorem shows that any type whose vertices are red and whose green edges comprise a spanning cycle will do.

Theorem 5.6. *Let $d \geq 2$ be an integer and let G be a connected d-regular graph of order n. Let τ_G be the type with vertex set $V(G)$ whose vertices are all red, with the edges of G coloured green and all other edges blue.*

If $p \leq 1/(d+2)$ then τ_G is p-core, and $\lambda_p(\tau_G) = q - \big(1 - p(d+2)\big)/n$.

Proof. Let $W = W_p(\tau)$ and let $x \in \Delta$. Let A be the adjacency matrix of G. Then

$$\begin{aligned} x^{\mathrm{t}}Wx &= q\Big(\sum_{v\in V(\tau)} x_v\Big)^2 - (q-p)\sum_{v\in V(\tau)} x_v^2 + px^{\mathrm{t}}Ax \\ &= q - (q-p-pd)\sum_{v\in V(\tau)} x_v^2 - p\Big(d\sum_{v\in V(\tau)} x_v^2 - x^{\mathrm{t}}Ax\Big) \\ &= q - \big(1-p(d+2)\big)\sum_{v\in V(\tau)} x_v^2 - p\sum_{uv\in E(G)} (x_u - x_v)^2. \end{aligned}$$

Since G is connected, both sums of squares are minimized only when $x_v = 1/n$ for all $v \in V(\tau)$. Because $1 - p(d+2) \geq 0$, the quantity $x^{\mathrm{t}}Wx$ has a unique maximum at this point. ∎

The preceding theorem does not preclude the type τ_G being p-core for values of p larger than $1/(d+2)$. For example, if $d = 2$ and G is an n-cycle, then Sidorenko's criterion (Proposition 3.21) can be called on to show that τ_G is p-core if and only if $p < 1/\big(2+2\cos(2\pi/n)\big)$; we omit the calculation.

Example 5.7. Let $H_r = K_{1,b}$ and $H_b = K_{b+1}$; thus $|H| = b+1$ and H consists of a green star $K_{1,b}$ with the remaining edges blue. As remarked in §3.6, in this case, $\kappa_p(H) = \mu_p(H)$ because missing edges in the extremal graphs can always be replaced by red edges. We observe that H is neither $\tau(0,1)$-colourable nor $\tau(b,0)$-colourable, and so $\kappa_p(H) \geq \max\{q, 1-q/b\}$.

Let τ be an optimal p-core type as in Theorem 3.25. Suppose $p > 1/2$. Any blue vertex of τ can be replaced by a red vertex without H becoming τ-colourable, and the p-value of τ has been increased. Since τ was optimal, it must be that the vertices were already red. So τ must have only red vertices, and by Theorem 3.23 these are joined by green edges. There cannot now be $b+1$ vertices, so $\tau \subset \tau(b,0)$ and $\kappa_p(H) \leq 1 - q/b$.

Suppose instead $p < 1/2$. If τ has a blue vertex v then it is joined to the others by green, by Theorem 3.23; if there is another then H is τ-colourable, so $\tau{=}\tau(0,1)$. Finally, if all the vertices are red, then these are joined to each other by blue and green edges, so no vertex can be joined to more than $b-1$ others by green. By Lemma 5.4, $\kappa_p(H) \leq 1 - q/b$.

We conclude that $\kappa_p(H) = \max\{q, 1-q/b\}$. A more general result with $H_r = K_a + \overline{K}_b$ and $H_b = \overline{H}_r$ is proved in [31].

Example 5.8. Let H_r be a star with perhaps an extra pendant edge added at one of its leaves, together with perhaps some isolated vertices. Let $H_b = \overline{H}_r$. Let $s = 2$ if H_r has an extra pendant edge and $s = 1$ otherwise; thus H_r contains s disjoint edges. Hence $\chi(H_b) \geq |H| - s$ and so H is not $\tau\big(|H|-s-1, 0\big)$-colourable. Certainly H is not $\tau(0,1)$-colourable because H_r is non-empty. Thus $\kappa_p(H) \geq \max\big\{q, 1 - q/\big(|H|-s-1\big)\big\}$.

On the other hand, H is $\tau(0,2)$-colourable because H_r is bipartite. H is also $\tau(1,1)$-colourable: to see that $H \subset m \times \tau(1,1)$ for large m, place the centre of the star plus its neighbour with the pendant edge (or any leaf if $s = 1$) in the red class and place all other vertices in the blue class.

Let τ be a p-core type for H with $\kappa_p(H) = \lambda_p(\tau)$. Suppose τ contains a blue vertex. If $|\tau| = 1$ then $\lambda_p(\tau) = q$. If $|\tau| > 1$ then, because neither $\tau(0,2) \subset \tau$ nor $\tau(1,1) \subset \tau$, there can be no green edge meeting a blue vertex. But then there can be no red vertices (by Theorem 3.23) and since the edges are red it must be that $p > 1/2$. Thus $\lambda_p(\tau) \leq p \leq 1 - q/\big(|H|-s-1\big)$.

On the other hand, if all vertices of τ are red, then they are joined by green and blue edges. If some vertex $u \in V(\tau)$ is joined to $|H|-s-1$ others by green, then H is contained in $m \times \tau$: place the star centre plus a leaf in the class V_u, if $s = 2$ place the neighbour with the pendant vertex together with

its pendant edge in some class $V_{u'}$ where uu' is green, and place all other vertices in distinct classes V_v where uv is green. So no vertex of τ is joined to $|H| - s - 1$ others by green and so $\lambda_p(\tau) \leq \max\{q, 1 - q/(|H| - s - 1)\}$ by Lemma 5.4.

In summary, $\kappa_p(H) = \max\{q, 1 - q/(|H| - 2)\}$ if H_r is a star plus isolated vertices, and $\kappa_p(H) = \max\{q, 1 - q/(|H| - 3)\}$ if H_r is a star with an extra pendant edge plus isolated vertices.

Example 5.9. We can now easily write down $\kappa_p(H)$ for all H with $H_b = \overline{H}_r$ and $|H| \leq 4$. The edit distance for these graphs was computed in [7] for $|H| \leq 3$, and for $|H| = 4$ in [5], with an alternative proof in [8].

If H is monochromatic then $\kappa_p(H)$ is given by Turán's theorem; for red H we have $\kappa_p(H) = q + p(1 - 1/(|H| - 1)) = 1 - p/(|H| - 1)$. In the other case when $|H| = 3$, there is a single edge in one colour, say red; this case is covered by Example 5.8, and $\kappa_p(H) = \max\{q, p\}$. Four other cases with $|H| = 4$ are covered by the same example; when the red edges comprise a star then $\kappa_p(H) = \max\{q, 1 - q/2\}$ and when H_r is a path of length three then $\kappa_p(H) = \max\{q, p\}$.

The final case is when $H_r = K_{2,2}$ and $H_b = \overline{H}_r = 2K_2$. Since H is not $\tau(1,1)$-colourable we have $\kappa_p(H) \geq 1 - pq$. Let τ be a p-core type for H as in Theorem 3.25. Then τ cannot have two blue vertices, since these must be joined by red or green and H would be τ-colourable. But H cannot have just red vertices, since either two are joined by green, in which case H is τ-colourable, or all are joined by blue, in which case $\lambda_p(H) \leq \max\{p, q\} < 1 - pq$. Thus τ has exactly one blue vertex, joined by green edges to the other (red) vertices. Now there cannot be two red vertices else these are joined by blue or green and H is τ-colourable. So $\tau = \tau(1,1)$ and $\kappa_p(H) = 1 - pq$.

For further examples, such as paths and cycles, and for more precise values of $\mathrm{kex}(H, n)$ in some instances, the reader is referred to [31].

5.2. Less simple examples

In the examples of §5.1, some type $\tau(a, b)$ whose edges are entirely green is always optimal. It is not immediately clear that there are graphs H which are not of this simple kind. The first example was provided by Balogh and Martin [8], who displayed a graph $H9$ of order 9 for which the edit distance

$1 - \min_p \kappa_p(H)$ is not realized by a type $\tau(a, b)$. We give a slightly smaller example which can be analysed fully.

Example 5.10. Let H_r be a six-cycle with a diagonal and let $H_b = \overline{H}_r$. We denote this graph H by C_6^*.

Let σ be the type with one blue and two red vertices, the red vertices being joined by a blue edge and the other edges being green. The vector assigning weight $1/(1+2p)$ to the blue vertex and $p/(1+2p)$ to each of the others shows $\lambda_p(\sigma) \geq 1 - p/(1+2p)$. The only subtype of σ with p-value greater than $1 - \min\{p, q\}$ is $\tau(1,1)$ with p-value $1 - pq$. Therefore σ is p-core for $0 < p < 1/2$, and moreover the vector just described is optimal by the converse of Fact 3.19 stated after Fact 3.20, so $\lambda_p(\sigma) = 1 - p/(1+2p)$.

Certainly C_6^* is not $\tau(2,0)$-colourable, because H_b is not bipartite. Consider a possible embedding of C_6^* in $m \times \sigma$. If we place a triangle of H_b in the blue class, there is a disjoint blue triangle which has to be placed amongst the two red classes, which is impossible. So at most two vertices can be placed in the blue class and, since no more than two can be placed in a red class, there must be exactly two in each class. The four vertices in the two red classes must comprise two red edges joined by a blue $K_{2,2}$. There is only one such subgraph in C_6^*, and the remaining two vertices span a red edge, which cannot be placed in the blue class. Therefore C_6^* is not σ-colourable. Hence $\kappa_p(C_6^*) \geq \max\{1 - q/2, 1 - p/(1+2p)\}$.

Now let τ be an optimal type for C_6^* as given by Theorem 3.25. Observe that C_6^* is $\tau(0,2)$-colourable because H_r is bipartite. Moreover C_6^* has three disjoint edges coloured red, red, blue and so it is $\tau(2,1)$-colourable. Notice too that C_6^* contains four vertices spanning a blue $K_{2,2}$ plus two red edges, with the other two vertices joined by red; thus τ cannot contain a red vertex joined by green to two other red vertices (these two must be joined by blue or green).

Suppose first that $p < 1/2$, so the edges of τ are green and blue by Theorem 3.23. Since $\tau(0,2) \not\subset \tau$, there is at most one blue vertex. If τ has a blue vertex it is joined by green to the red vertices, and there can be at most two of these (else C_6^* would be τ-colourable because red vertices are joined by blue or green). Because $\tau(2,1) \not\subset \tau$, if there are two red vertices they must be joined by blue; thus $\tau \subset \sigma$. Suppose on the other hand that there are no blue vertices. We know that a red vertex can be joined by green to at most one other red vertex. Then Lemma 5.4 shows that $\lambda_p(\tau) \leq \max\{q, 1 - q/2\} < 1 - p/(1+2p)$.

Suppose now that $p \geq 1/2$. The edges of τ are green or red. Since C_6^* is both $\tau(2,1)$-colourable and $\tau(3,0)$-colourable (because H_b is 3-chromatic), if τ has two red vertices it can have no other vertices so $\tau = \tau(2,0)$. If τ has only blue vertices, they must be joined by red edges because $\tau(0,2) \not\subset \tau$, but then $\lambda_p(\tau) \leq p \leq 1 - q/2$. Thus τ has exactly one red vertex. There must be at least two blue vertices else $\lambda_p(\tau) \leq \lambda_p\big(\tau(1,1)\big) = 1 - pq \leq 1 - q/2$. So τ contains a red vertex joined to two blue vertices by green, and the blue vertices are joined by a red edge. But C_6^* contains four vertices spanning a red $K_{2,2}$ plus two blue edges, with the other two vertices joined by red; so in this case C_6^* is τ-colourable.

We conclude that $\kappa_p(C_6^*) = \max\big\{1 - q/2, 1 - p/(1+2p)\big\}$, and that when $p \neq 0, \frac{1}{2}, 1$ then σ and $\tau(2,0)$ are the only optimal types.

As noted in §3.6, C_6^* gives an example of a graph for which $\kappa_p \neq \mu_p$, because a complete 5-partite graph coloured green does not contain C_6^*, and therefore $\mu_p(C_6^*) \geq 4/5$ for all p. Hence $\mu_p(C_6^*) > \kappa_p(C_6^*)$ for $1/3 < p < 3/5$.

The next example requires less work and will be helpful in §5.4. It is based on the following lemma about green triangles, which is itself straightforward and useful.

Lemma 5.11. *Let τ be a type whose vertices are all red and which contains no green triangle. If $p \leq 1/2$ then $\lambda_p(\tau) \leq (p-q)/|\tau| + q + p/2 \leq 1 - p/2$.*

Proof. Consider the graph of order n in $\mathcal{Q}(\tau)$ of maximum weight, as in Remark 3.6. It has no green triangle so by Turán's theorem it has at most $n^2/4$ green edges. So if x is a vector maximizing $x^{\mathrm{t}} W_p(\tau) x$ then $\lambda_p(\tau)\binom{n}{2} = w_p(G) + O(n) \leq \big((p-q)\sum_u x_u^2 + q + p/2\big)\binom{n}{2} + O(n)$, and hence $\lambda_p(\tau) \leq (p-q)\sum_u x_u^2 + q + p/2$. Since $\sum_u x_u^2 \geq 1/|\tau|$ and $p - q < 0$ we obtain $\lambda_p(\tau) \leq (p-q)/|\tau| + q + p/2$. ∎

Example 5.12. Let H_1 be the graph of Example 5.7, namely, a green star $K_{1,b}$ with the remaining edges blue. Let H_2 be a green K_3 and let $\mathcal{H} = \{H_1, H_2\}$. Neither graph in $\mathcal{H}$ is $\tau(0,1)$-colourable or $\tau(2,0)$-colourable. Nor is either σ-colourable, where σ is the type of order $2(b-1)$ with red vertices, whose green edges form $K_{b-1,b-1}$ and whose remaining edges are blue. The uniform weighting shows that $\lambda_p(\sigma) \geq q + p/2 + (p-q)/2(b-1)$. Therefore $\kappa_p(\mathcal{H}) \geq \max\big\{q, 1 - q/2, q + p/2 + (p-q)/2(b-1)\big\}$.

Let τ be an optimal p-core type for $\mathcal{H}$. If $p > 1/2$ then any blue vertex of τ can be re-coloured red without H_1 or H_2 becoming τ-colourable, and this would increase $\lambda_p(\tau)$. Therefore all vertices of τ must be red and joined

by green edges, so $\tau \subset \tau(2,0)$. Thus $\kappa_p(\mathcal{H}) \leq 1 - q/2$, and the same holds for $p = 1/2$ by Theorem 3.1. So suppose $p < 1/2$. If τ has a blue vertex v then, as in Example 5.7, there can be no other vertex, so $\tau{=}\tau(0,1)$ and $\lambda_p(\tau) = q$. This leaves the case where τ has only red vertices, joined by blue or green edges. As in the proof of Lemma 5.4, the equation $\lambda_p(\tau) \leq q + \big(p(m+1) - q\big)/|\tau|$ holds, where m is the maximum number of green edges meeting a vertex. Clearly $m \leq b-1$ because H_1 is not τ-colourable. If $p(m+1) - q < 0$ we have $\lambda_p(\tau) \leq q$, and otherwise, if $|\tau| \geq 2(b-1)$, we have $\lambda_p(\tau) \leq q + \big(p(m+1) - q\big)/2(b-1) \leq q + p/2 + (p-q)/2(b-1)$. By Lemma 5.11 this inequality still holds if $|\tau| \leq 2(b-1)$, bearing in mind that $p - q < 0$.

We conclude that $\kappa_p(\mathcal{H}) = \max\big\{q, 1-q/2, q+p/2+(p-q)/2(b-1)\big\}$.

Remark 5.13. Example 5.12 shows that the strict inequality $\kappa_p\big(\{H_1,H_2\}\big) < \min\big\{\kappa_p(H_1), \kappa_p(H_2)\big\}$ is possible, because $\kappa_p(H_1) = \max\{q, 1-q/b\}$ by Example 5.7, and $\kappa_p(H_2) = 1/2 + \max\{p/2, q/2\}$ by Turán's theorem. Thus the inequality is strict for $1/(b+1) < p < 1/2$.

5.3. Complete bipartite examples

In what remains, we study H with H_r complete bipartite and $H_b = \overline{H}_r$. Examples 5.5 and 5.9 show that the cases $H_r = K_{1,t}$ and $H_r = K_{2,2}$ can be treated easily, but even $H_r = K_{2,t}$ becomes much more involved.

We first prove a general fact about optimal types for these graphs.

Lemma 5.14. *Let $H_r = K_{s,t}$ where $s \leq t$ and let $H_b = \overline{H}_r$. Let τ be a p-core type such that H is not τ-colourable and $\kappa_p(H) = \lambda_p(\tau)$, as guaranteed by Theorem 3.25. Then either $\tau = \tau(s-1,1)$ and $\kappa_p(H) = 1 - pq/\big(1+p(s-2)\big)$, or all the vertices of τ are red.*

If $p > \frac{1}{2}$, then $\tau = \tau(s-1,1)$ if $t = s$ and $\tau = \tau(t-1,0)$ otherwise.

Proof. We use Theorem 3.23 repeatedly. There cannot be two blue vertices in τ, for they would be joined by a red or a green edge and then H would be τ-colourable. If τ has exactly one blue vertex, then it is joined to the other vertices by green edges. The red vertices are joined to each other by blue or green edges, so if there are s of them then H would be τ-colourable (by placing t vertices in the blue class of $m \times \tau$ and the other s vertices in distinct red classes). Thus τ has at most $s-1$ red vertices and so $\tau \subset \tau(s-1,1)$.

Since H is not $\tau(s-1,1)$-colourable, it must be that $\tau = \tau(s-1,1)$ and $\kappa_p(H) = \lambda_p(\tau) = 1 - pq/\big(1+p(s-2)\big)$.

If $p > \frac{1}{2}$ and $\tau \neq \tau(s-1,1)$ then all the vertices are red and all the edges are green. Since H is $\tau(t,0)$-colourable but not $\tau(t-1,0)$-colourable, it must be that $\tau = \tau(t-1,0)$. Now $\lambda_p\big(\tau(t-1,0)\big) = 1 - q/(t-1)$, which is larger than $\lambda_p\big(\tau(s-1,1)\big)$ if $p > 1/(t-s+1)$. ∎

Given an ordinary uncoloured graph G, let τ_G be the type with vertex set $V(G)$ whose vertices are all red, with the edges of G coloured green and all other edges blue. By taking the vector $x = \big(1/|G|, \dots, 1/|G|\big)$ we have

$$\lambda_p(\tau_G) \geq x^{\mathrm{t}} W_p(\tau_G) x = q + \frac{p-q}{|G|} + 2p\frac{e(G)}{|G|^2} = q + \frac{p(d+2)-1}{|G|}$$

where $e(G)$ is the number of edges of G and d is the average degree.

By Lemma 5.14, the determination of $\kappa_p(H)$ comes down to finding those G with $\lambda_p(\tau_G) > 1 - pq/\big(1+p(s-2)\big)$ for which H is not τ_G-colourable, if indeed there are any. But this task is not straightforward.

Remark 5.15. If H is τ_G-colourable, that is, $H \subset m \times \tau_G$ for some m, then the vertices of the blue K_s lie in different classes, as do the vertices of the blue K_t. So there is some number c, $0 \leq c \leq s$ of classes, which contain both a vertex of K_s and a vertex of K_t, and the remaining $s+t-2c$ vertices are in distinct classes. The c classes that contain two vertices must be joined by green to all the other $s+t-c-1$ other classes. This implies that G contains $K_c + K_{s-c,t-c}$ as a subgraph. Notice that $K_{s-1} + K_{1,t-s+1}$ contains $K_s + K_{0,t-s} = K_s + \overline{K}_{t-s}$, and if $s = t$ then so does $K_{s-2} + K_{2,t-s+2}$. Therefore, the condition that H be not τ_G-colourable is that none of $K_s + \overline{K}_{t-s}$ and $K_c + K_{s-c,t-c}$, $0 \leq c \leq s-2$, are subgraphs of G, though the case $c = s-2$ is redundant if $s = t$.

Example 5.16. Let $H_r = K_{2,t}$ and $H_b = \overline{H}_r$. We know from Example 5.9 that $\tau(1,1)$ is optimal for all p if $t = 2$, and $\kappa_p(H) = \lambda_p\big(\tau(1,1)\big) = 1 - pq = q + p^2$. However, this is not true for any $t \geq 4$. If G is a graph containing neither $K_{2,t}$ nor $K_2 + \overline{K}_{t-2}$ then H is not τ_G-colourable, by Remark 5.15. Take such a graph G which is large and reasonably dense; then τ_G will be better than $\tau(1,1)$ for some small values of p. In fact, the graph $G = K_{3,3}$ already gives some information. We have $|G| = 6$ and $d = 3$ so $\lambda_p(\tau_G) \geq q + (5p-1)/6 > q + p^2$ if $1/3 < p < 1/2$. So, if $t \geq 4$, $\tau(1,1)$ is non-optimal for $1/3 < p < 1/2$.

Example 5.17. Let $H_r = K_{s,s}$ and $H_b = \overline{H}_r$, where $s \geq 2$. Let $G = K_{2s-2} - (s-1)K_2$, that is, G is a $K_{s-1,s-1}$ with a 1-factor removed. Then G doesn't contain K_s or $K_c + K_{s-c,s-c}$ for any c, $0 \leq c \leq s-3$, and so H is not τ_G-colourable. We have $|G| = 2s-2$ and $d = 2s-4$ so $\lambda_p(\tau_G) \geq 1 - 1/(2s-2)$. Hence $\lambda_p(\tau_G) > \lambda_p\big(\tau(s-1,1)\big)$ if $p^2(2s-2) - sp + 1 < 0$. Since the roots of this quadratic lie between zero and one, there is a value of p for which the inequality holds provided $s^2 - 8(s-1) > 0$, or $s \geq 7$.

The meaning of this for edit distance (see §2.2) is that, for all $s \geq 7$, the edit distance for $\mathrm{Forb}(K_{s,s})$, namely $1 - \min_p \kappa_p(H)$, is realized by a type τ not of the form $\tau(a,b)$, because by Lemma 5.14 a type with only green edges would have to be either $\tau(s-1,1)$ or $\tau(s-1,0)$. Clearly $\lambda_p\big(\tau(s-1,0)\big) \leq \lambda_p\big(\tau(s-1,1)\big)$ for all p, and the minimum value of $\lambda_p\big(\tau(s-1,1)\big)$ is less than $1 - 1/(2s-2) = \lambda_p(\tau_G)$.

Another choice of G which works here, and which is a slightly better choice, is $G = K_{2s-1} - (s-2)K_2 - P_3$, where P_3 is a path of length two disjoint from the missing edges.

5.4. $K_{3,3}$

The case $H_r = K_{3,3}$ and $H_b = \overline{H}_r$ is a particularly interesting one. It appears as a special example in the conjectures of Richer (Conjecture 4.2) and of Diwan and Mubayi (in §4), and it was mentioned by Alon and Stav [5] as the smallest outstanding case whose edit distance they had not computed. The edit distance was subsequently shown by Balogh and Martin [8] to be $3 - 2\sqrt{2}$, this being $1 - \min_p \lambda_p\big(\tau(2,1)\big)$, minimized at $p = \sqrt{2} - 1$.

As shown by Lemma 5.14, $\kappa_p(H) = \lambda_p\big(\tau(2,1)\big)$ for $p > 1/2$, and this raises the question whether this equation holds for all p. In fact, it is relatively easy to show that it holds for $p > 1/3$ (Example 5.18), which is enough to evaluate the edit distance, so giving an alternative and somewhat simpler proof than that in [8]. It turns out that the equation does not hold for small p (Example 5.19), though the argument is a little delicate. After giving these examples we shall examine for which values of p the equation can be shown to still hold.

Example 5.18. Let $H_r = K_{3,3}$ and $H_b = \overline{H}_r$. Let $1/3 < p < 1/2$ and let τ be a p-core type for which H is not τ-colourable and with $\kappa_p(H) = \lambda_p(\tau)$, as in Theorem 3.25. Then $\lambda_p(\tau) \geq \lambda_p\big(\tau(2,1)\big) = 1 - pq/(1+p) > 1 - p/2$. Lemma 5.14 shows that either $\tau = \tau(2,1)$ or all the vertices of τ are red,

but the latter is impossible by Lemma 5.11, for then τ would contain $\tau(3,0)$ and H is $\tau(3,0)$-colourable.

The minimum of $\lambda_p\big(\tau(2,1)\big)$ occurs at $p = \sqrt{2}-1 > 1/3$, and so $\kappa_p(H)$ has a local minimum at this point. By Theorem 3.1 this local minimum is also a global minimum, and so the edit distance of $K_{3,3}$ is $3-2\sqrt{2}$.

Example 5.19. Let $H_r = K_{3,3}$ and $H_b = \overline{H}_r$. Let G be a graph that contains neither K_3 nor $K_{3,3}$. Then H is not τ_G-colourable by Remark 5.15. As pointed out earlier, $\lambda_p(\tau_G) \geq q + \big(p(d+2)-1\big)/|G|$, whereas $\lambda_p\big(\tau(2,1)\big) = 1 - pq/(1+p) = q + 2p^2/(1+p)$. Thus $\lambda_p(\tau_G) > \lambda_p\big(\tau(2,1)\big)$ if $(p+1)\big(p(d+2)-1\big) > 2p^2|G|$; in order for the graph G to be useful there needs to be a value of p for such that this inequality holds, and this requires $(d+3)^2 > 8|G|$.

Finding a graph G with these properties is not too easy; the constant 8 is just too large for random constructions or for standard pseudo-random constructions with d around $|G|^{1/2}$. But the beautiful construction of bipartite $K_{3,3}$-free graphs by Brown [14] does work. For each odd prime power r, the construction gives a graph G with $|G| = 2r^3$ and which is (r^2-r)-regular. The required inequality $(d+3)^2 > 8|G|$ holds for $r \geq 19$. When $r = 19$, the inequality $\lambda_p(\tau_G) > \lambda_p\big(\tau(2,1)\big)$ holds in the range $1/219 \leq p \leq 1/124$. For each $r \geq 23$, it can be checked that the inequality holds at $p = 1/12r$ and at $p = 1/6r$, and thus for all p between these values. By Bertrand's postulate this means that, for every $p \leq 1/124$, there is some r for which Brown's graph G satisfies $\kappa_p(H) \geq \lambda_p(\tau_G) > \lambda_p\big(\tau(2,1)\big)$.

In the light of the preceding discussion, it would be interesting to know for which values of p it is true that $\kappa_p(H) = \lambda_p\big(\tau(2,1)\big)$. We describe an approach for reducing the bound $p \geq 1/3$ of Example 5.18, which could be applied to other extremal problems as well.

We suppose from now on that $p \leq 1/3$ and that $\tau(2,1)$ is not optimal; that is, there is another p-core type τ which is optimal, and $\lambda_p(\tau) > \lambda_p\big(\tau(2,1)\big) = 1 - pq/(1+p)$. By Lemma 5.14, all the vertices of τ are red, and its edges are green and blue.

Consider a sequence of graphs (G^n) of maximal weight in $\mathcal{Q}(\tau)$, as described in Remark 3.6. We have $\liminf \delta_p(G^n)/n = \lambda_p(\tau) > 1-pq/(1+p)$. Notice that the largest blue complete subgraph in any graph G^n has at most $|\tau|$ vertices, because all the vertices of τ are red. In particular, the type $\tau(0,1)$, consisting of a single blue vertex, is not contained in (G^n), and so any type that is contained in (G^n) can have only red vertices. Moreover,

by Remark 5.15, no type contained in (G^n) contains a green triangle or a green $K_{3,3}$.

We define two special types, $F(t)$ and $G(k)$. All the vertices of both types are red. The type $F(t)$ has $t+1$ vertices, and contains a green star $K_{1,t}$, all other edges being blue. Thus $F(t)$ is the graph H_1 of Example 5.12 with $b = t$ and all vertices coloured red. The graph $G(k)$ has $k+2$ vertices, and contains a green $K_{2,k}$, all other edges being blue. For convenience we label the vertices of $F(t)$ with $1, \ldots, t+1$ so that the green edges meet vertex 1. Similarly, we label $G(k)$ with $1, \ldots, k+2$ so the green edges meet vertices 1 and 2.

If (G^n) contains $F(t)$ or $G(k)$ we apply Lemma 3.11 to obtain an extension σ. By choosing the vector x carefully, we can arrange that the new vertex in the extension is joined to the rest of the type in a way that is useful to us. We consider the case that (G^n) contains $G(k)$ first because, though the case $F(t)$ is easier, it makes use of the case $G(k)$.

Claim 1. *(G^n) does not contain $G(k)$ for $k \geq 1/p - 3$.*

Proof. If the claim fails then, for any vector x of our choosing, Lemma 3.11 gives an extension σ of $G(k)$ contained in (G^n), with $\sum_{j=1}^{k+2} x_j w(jv) > 1-pq/(1+p)$. As observed above, the vertices of σ must all be red. We call a vertex $j \in \{1, \ldots, k+2\}$ a *green neighbour* if vj is green. Since (G^n) contains σ, σ cannot contain a green triangle or a green $K_{3,3}$. In particular, the sets $\{1,2\}$ and $\{3, \ldots, k+2\}$ cannot both contain a green neighbour. We choose a number $0 \leq \alpha \leq 1/k$ and put $x_j = \alpha$ for $j \geq 3$, $x_1 = x_2 = (1 - k\alpha)/2$, so $x \in \Delta$; the exact choice of α remains to be decided.

First note that we must have a green neighbour, otherwise $\sum_{j=1}^{k+2} x_j w(jv) \leq q \leq 1 - pq/(1+p)$. Suppose the event occurs that the number of green neighbours in $\{3, \ldots, k+2\}$ is either one or two. Then $\sum_{j=1}^{k+2} x_j w(jv) \leq q(1-k\alpha) + 2\alpha + (k-2)q\alpha$. So we can forbid this event by choosing α with $q(1-k\alpha)+2\alpha+(k-2)q\alpha \leq 1-pq/(1+p)$; that is, $\alpha \leq p/(1+p)$. The event that $\{3, \ldots, k+2\}$ contains at least three green neighbours means that at least one of $v1$ or $v2$ is red, else σ would contain a green $K_{3,3}$, so we can forbid this event too by choosing $(p+q)(1-k\alpha)/2 + k\alpha \leq 1 - pq/(1+p)$, or $k\alpha \leq 1 - 2pq/(1+p)$. Thus choosing α to satisfy both these upper bounds ensures that $\{3, \ldots, k+2\}$ contains no green neighbours.

We can forbid the event that exactly one of the vertices in $\{1,2\}$ is a green neighbour by choosing α with $(1+q)(1-k\alpha)/2 + kq\alpha \leq 1-pq/(1+p)$, or $k\alpha \geq (1-3p)/(1+p)$. If we can further choose α so that $(1-k\alpha) +$

$(k-1)q\alpha + p\alpha \leq 1 - pq/(1+p)$, or $(1+(k-2)p)\alpha \geq pq/(1+p)$, then we can also forbid the event that both 1 and 2 are green neighbours and some vj, $j \in \{3,\ldots,k+2\}$, is red. Therefore a choice of α that satisfies both of these lower bounds as well will guarantee that $\sigma = G(k+1)$.

Finally, if we can choose an α that also satisfies the additional bound $(1-k\alpha) + kq\alpha \leq 1 - pq/(1+p)$, or $k\alpha \geq q/(1+p)$, then we obtain a contradiction; the inequality $\sum_{j=1}^{k+2} x_j w(jv) > 1 - pq/(1+p)$ becomes impossible.

If $1/p - 3 \leq k \leq 1/p - 1$ we choose $\alpha = p/(1+p) \leq 1/(k+2)$. Then $k\alpha \leq (1-p)/(1+p) \leq 1 - 2pq/(1+p)$, so the upper bounds for α are all satisfied. The first lower bound holds because $kp \geq 1 - 3p$, and so does the second because $1 + (k-2)p \geq q$. So if $1/p - 3 \leq k \leq 1/p - 1$ and (G^n) contains $G(k)$, then it also contains $G(k+1)$.

If $k \geq 1/p - 1$ we choose $\alpha = q/k(1+p) < 1/k$. Now $q/k \leq p$ and $k\alpha = (1-p)/(1+p)$ so once again the upper bounds hold. The first lower bound holds because $q \geq (1-3p)$, as does the second because $(1+(k-2)p) \geq kp$. Moreover, since $k\alpha = q/(1+p)$ the additional bound is satisfied, and so this time we obtain a contradiction.

Therefore, if (G^n) contains $G(k)$ and $k \geq 1/p - 3$ then, by noting that $G(k)$ contains $G(k-1)$ and reducing k if necessary, we may presume that $k \leq 1/p - 2$. Then (G^n) contains $G(k+1)$ and, since $k+1 \leq 1/p - 1$, (G^n) also contains $G(k+2)$. But $k+2 \geq 1/p - 1$ and we now obtain a contradiction, so proving the claim. ■

We now look further at what happens if (G^n) contains $F(t)$. We proceed in a similar manner to the discussion for $G(k)$, though the argument is much simpler.

Claim 2. *If (G^n) contains $F(t)$ then $6 + \lfloor 2(t-1)p/q \rfloor < 1/p$.*

Proof. If (G^n) contains $F(t)$ then it contains an extension σ of $F(t)$ with $\sum_{j=1}^{t+1} x_j w(jv) > 1 - pq/(1+p)$. Choosing $0 \leq \alpha \leq 1/t$, we set $x_1 = (1-t\alpha)$ and $x_j = \alpha$ for $j \geq 2$. We prevent $v1$ being red by choosing α so that $p(1-t\alpha) + t\alpha \leq 1 - pq/(1+q)$, or $t\alpha \leq 1/(1+p)$. If in fact $v1$ is green, then none of vj, $j \geq 2$ is green since σ contains no green triangle, and we can guarantee that these edges are all blue by choosing α so that $(1-t\alpha) + q(t-1)\alpha + p\alpha \leq 1 - pq/(1+q)$, or $\big(p(t-1)+q\big)\alpha \geq pq/(1+p)$. This means that if $v1$ is green then σ contains $F(t+1)$. On the other hand, if $v1$ is blue, we can ensure that at least k of the edges vj, $j \geq 2$ are green

by choosing α so that $q(1-t\alpha)+(k-1)\alpha+(t-k+1)q\alpha \le 1-pq/(1+q)$, or $(k-1)\alpha \le 2p/(1+p)$. This last condition means that if $v1$ is blue then σ contains $G(k)$.

We choose $\alpha = pq/(1+p)\big(p(t-1)+q\big)$, so $\alpha < p/\big(p(t-1)+q\big) < 1/t$ and $t\alpha < tp/(1+p)\big(p(t-1)+q\big) < 1/(1+p)$, and therefore the first upper bound is satisfied. The lower bound holds by definition of α. Finally, the last upper bound holds provided $(k-1)q/\big(p(t-1)+q\big) \le 2$, or $k \le 3+2(t-1)p/q$.

Suppose now that (G^n) contains $F(t)$ but does not contain $G(k)$, where $k = 3+\lfloor 2(t-1)p/q \rfloor$. Then (G^n) contains $F(t+1)$. But then it must likewise contain $F(t+2)$, and $F(t+3)$, and in fact $F(n)$ for all n. But then (G^n) contains arbitrarily large blue cliques, contradicting the fact noted above that the largest blue clique can have at most $|\tau|$ vertices.

Hence if (G^n) contains $F(t)$ for some t, then it contains $G(k)$ where $k = 3+\lfloor 2(t-1)p/q \rfloor$. But, by Claim 1, (G^n) does not contain $G(k)$ if $k \ge 1/p-3$. We conclude that if (G^n) contains $F(t)$ then $6+\lfloor 2(t-1)p/q \rfloor < 1/p$. ■

Recall now Example 5.12. Taking $b = 10$ we have $\kappa_p(\mathcal{H}) = \max\{q, 1-q/2, q+p/2+(p-q)/18\}$. Certainly $\lambda_p(\tau) > 1-pq/(1+p) > 1-p = q$ and $\lambda_p(\tau) > 1-q/2$. Moreover $\lambda_p(\tau) - \big[q+p/2+(p-q)/18\big] > 1-pq/(1+p) - \big[q+p/2+(p-q)/18\big] = (5p-1)^2/18(1+p) \ge 0$. Therefore $\lambda_p(\tau) > \kappa_p(\mathcal{H})$. Now (G^n) is a sequence of graphs in $\mathcal{Q}(\tau)$ of maximum weight, and so $w_p(G^n) = \big(\lambda_p(\tau)+o(1)\big)\binom{|G|}{2}$ by Remark 3.6. It follows that the graphs (G^n) contain a member of $\mathcal{H}$ if n is large. Now τ has only red vertices and does not contain a green triangle. So, since G^n is in $\mathcal{Q}(\tau)$, it cannot contain H_2, which is a green triangle. Therefore G^n must contain H_1. The fact that $G^n \in \mathcal{Q}(\tau)$ now means that τ contains $F(10)$.

So we have a contradiction to Claim 2 if $6+\lfloor 18p/q \rfloor \ge 1/p$. This condition holds when $p = 1/8$, and so it holds for all larger values of p. The contradiction means that $\tau(2,1)$ is extremal for H; thus we have reduced the bound of Example 5.18 to $p \ge 1/8$.

We can extend this result to the interval $1/9 \le p < 1/8$ with a further small observation. It is enough to show that (G^n) contains either $G(6)$ or $F(13)$, because $6+\lfloor 24p/q \rfloor \ge 9 \ge 1/p$. Assume, to the contrary, that τ contains neither $G(6)$ nor $F(13)$. Now τ has only red vertices and no vertex is joined to more than 12 others by green, since $F(13) \not\subset \tau$. The simple argument of Lemma 5.4 shows that $\lambda_p(\tau) \le q+(13p-q)/|\tau|$. Thus $1-pq/(1+p) < q+(13p-q)/|\tau|$, or $|\tau| < (1+p)(14p-1)/2p^2 \le (1+1/8)(14/8-1)9^2/2 < 35$.

Let $x \in \Delta$ be such that $x^t W_p(\tau)x = \lambda_p(\tau)$. Given a vertex $u \in V(\tau)$, let $N(u) = \{v \in V(\tau) : uv \text{ is green}\}$ and let $d(u) = \sum \{x_j : j \in N(u)\}$. By Fact 3.19, $\lambda_p(\tau) = (p-q)x_u + q + pd(u)$, so $d(u) = (\lambda_p(\tau) - q)/p + (q/p - 1)x_u$. Now $\lambda_p(\tau) - q > 1 - pq/(1+p) - q = 2p^2/(1+p)$, and so $(\lambda_p(\tau) - q)/p > 2p/(1+p) \geq 1/5$. Thus $d(u) > 1/5 + (q/p - 1)x_u \geq 1/5 + 6x_u$ holds for all $u \in V(\tau)$.

Choose u with x_u maximal. In particular, $x_u \geq 1/|\tau| > 1/35$. Now $N(u)$ has at most 12 vertices so we can choose $v, w \in N(u)$ with $x_v + x_w \geq 2d(u)/12 = d(u)/6$. Since τ contains no green triangle we have $N(u) \cap N(v) = N(u) \cap N(w) = \emptyset$. So, writing $d = \sum \{x_j : j \in N(v) \cap N(w)\}$ we have $d \geq d(v) + d(w) - (1 - d(u)) \geq -2/5 + 6x_u + 6(x_v + x_w) \geq -2/5 + 6x_u + d(u) \geq -1/5 + 12x_u$. But $N(v) \cap N(w)$ contains at most 5 vertices because τ contains no $\tau_{G(6)}$. Therefore $d \leq 5x_u$, giving $12x_u - 1/5 \leq 5x_u$ or $x_u \leq 1/35$, contradicting $x_u > 1/35$.

In conclusion, we have shown that $\kappa_p(\tau(H)) = \lambda_p(\tau(2,1))$ for $p \geq 1/9$.

Acknowledgement. The authors are grateful to Ryan Martin for stimulating discussions and to Miklós Simonovits for pointing them towards the literature on the extremal theory of directed graphs and multigraphs.

References

[1] V. E. Alekseev, Hereditary classes and coding of graphs, *Probl. Cybern.*, **39** (1982) 151–164 (in Russian).

[2] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, *Combinatorica*, **20** (2000), 451–476.

[3] N. Alon and A. Shapira, A characterization of the (natural) graph properties testable with one-sided error, *Proc. 46th IEEE Symp. Foundations of Computer Science* (2005), 429–438.

[4] N. Alon and U. Stav, What is the furthest graph from a hereditary property? *Random Structures and Algorithms*, **33** (2008), 87–104.

[5] N. Alon and U. Stav, The maximum edit distance from hereditary graph properties, *J. Combinatorial Theory (Ser. B)*, **98** (2008), 672–697.

[6] N. Alon and U. Stav, Stability type results for hereditary properties, (preprint).

[7] M. Axenovich, A. Kézdy and R. Martin, On the editing distance of graphs, *J. Graph Theory*, **58** (2008), 123–138.

[8] J. Balogh and R. Martin, Edit distance and its computation, *Electronic Journal of Combinatorics*, **15** (2008), R20.

[9] J. Beck, Combinatorial games.Tic-tac-toe theory. *Encyclopedia of Mathematics and its Applications,* **114**. Cambridge University Press, Cambridge (2008). xiv+732pp.

[10] B. Bollobás, *Extremal Graph Theory,* Academic Press, London (1978) xx+488pp.

[11] B. Bollobás and A. Thomason, Hereditary and monotone properties of graphs, in "The Mathematics of Paul Erdős II" (R. L. Graham and J. Nešetřil, eds.) *Algorithms and Combinatorics,* **14** Springer-Verlag (1997), 70–78.

[12] B. Bollobás and A. Thomason, The structure of hereditary properties and colourings of random graphs, *Combinatorica,* **20** (2000), 173–202.

[13] J. A. Bondy and Zs. Tuza, A weighted generalization of Turán's theorem, *J. Graph Theory,* **25** (1997), 267–275.

[14] W. G. Brown, On graphs that do not contain a Thomsen graph, *Canad. Math. Bull.,* **9** (1966), 281–285.

[15] W. G. Brown, P. Erdős and M. Simonovits, Extremal problems for directed graphs, *J. Combinatorial Theory (Ser. B),* **15** (1973), 77–93.

[16] W. G. Brown, P. Erdős and M. Simonovits, Inverse extremal digraph problems, "Finite and Infinite Sets, Eger (Hungary) 1981", *Colloq. Math. Soc. János Bolyai,* **37**, Akad. Kiadó, Budapest (1985), 119–156.

[17] W. G. Brown, P. Erdős and M. Simonovits, Algorithmic solution of extremal digraph problems, *Trans. Amer. Math. Soc.,* **292** (1985), 421–449.

[18] W. G. Brown and M. Simonovits, Extremal multigraph and digraph problems, "Paul Erdős and his mathematics, II (Budapest, 1999)" *Bolyai Soc. Math. Stud.,* **11**, János Bolyai Math. Soc., Budapest (2002), 157–203.

[19] D. De Caen and Z. Füredi, The maximum size of 3-uniform hypergraphs not containing a Fano plane, *J. Combinatorial Theory (Ser. B),* **78** (2000), 274–276.

[20] A. Diwan and D. Mubayi, Turán's theorem with colors, (preprint).

[21] P. Erdős, Some recent results on extremal problems in graph theory (Results), *Theory of Graphs* (Internat. Sympos., Rome, 1966), Gordon and Breach, New York; Dunod, Paris (1967), 117–123.

[22] P. Erdős, On some new inequalities concerning extremal properties of graphs. *Theory of Graphs* (Proc. Colloq. Tihany, 1966), Academic Press, New York (1968), 77–81.

[23] P. Erdős and M. Simonovits, A limit theorem in graph theory, *Studia Sci. Math. Hungar.,* **1** (1966), 51–57.

[24] P. Erdős and A.H. Stone, On the structure of linear graphs, *Bull. Amer. Math. Soc.,* **52** (1946), 1087–1091.

[25] Z. Füredi and A. Kündgen, Turán problems for integer-weighted graphs, *J. Graph Theory,* **40** (2002), 195–225.

[26] Z. Füredi and M. Simonovits, Triple systems not containing a Fano configuration, *Combinatorics, Probability and Computing,* **14** (2005), 467–484.

[27] Gy. Katona, T. Nemetz and M. Simonovits, On a problem of Turán in the theory of graphs (in Hungarian), *Mat. Lapok,* **15** (1964) 228–238.

[28] P. Keevash and B. Sudakov, The Túran number of the Fano plane, *Combinatorica,* **25** (2006), 561–574.

[29] J. Komlós and M. Simonovits, Szemerédi's regularity lemma and its applications in graph theory, *Combinatorics, Paul Erdős is eighty,* Vol. 2 (D. Miklós, V. T-Sós and T. Szönyi, eds.), János Bolyai Math. Soc., Budapest (1996), 295–352.

[30] L. Lovász and B. Szegedy, Testing properties of graphs and functions (in press).

[31] E. Marchant, (in preparation).

[32] E. Marchant and A. Thomason, The structure of hereditary properties and 2-coloured multigraphs, (submitted).

[33] H.-J. Prömel and A. Steger, Excluding induced subgraphs III: a general asymptotic, *Rand. Struct. Alg.,* **3** (1992), 19–31.

[34] D. C. Richer, Ph. D. thesis, University of Cambridge (2000).

[35] V. Rödl and A. Sidorenko, On the jumping constant conjecture for multigraphs, *J. Combinatorial Theory (Ser. A),* **69** (1995), 347–357.

[36] A. F. Sidorenko, Boundedness of optimal matrices in extremal multigraph and digraph problems, *Combinatorica,* **13** (1993), 109–120.

[37] M. Simonovits, A method for solving extremal problems in graph theory, stability problems, *Theory of Graphs* (Proc. Colloq. Tihany, 1966), Academic Press, New York (1968), 279–319.

[38] M. Simonovits, Extremal graph theory, *Selected topics in graph theory* **2** (L. W. Beineke and R. J. Wilson, eds.) Academic Press, London (1983), 161–200.

Edward Marchant
DPMMS
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WB
UK
e-mail:
E.Marchant@dpmms.cam.ac.uk

Andrew Thomason
DPMMS
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WB
UK
e-mail:
A.G.Thomason@dpmms.cam.ac.uk

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 287–325.

Regularity Lemmas for Graphs

VOJTĚCH RÖDL* and MATHIAS SCHACHT

Szemerédi's regularity lemma proved to be a fundamental result in modern graph theory. It had a number of important applications and is a widely used tool in extremal combinatorics. For some further applications variants of the regularity lemma were considered. Here we discuss several of those variants and their relation to each other.

1. Introduction

Szemerédi's regularity lemma is one of the most important tools in extremal graph theory. It has many applications not only in graph theory, but also in combinatorial number theory, discrete geometry, and theoretical computer science. The first form of the lemma was invented by Szemerédi [47] as a tool for the resolution of a famous conjecture of Erdős and Turán [9] stating that any sequence of integers with positive upper density must contain arithmetic progressions of any finite length.

The regularity lemma roughly states that every graph may be approximated by a union of induced random-like (quasi-random) bipartite subgraphs. Since the quasi-randomness brings important additional information, the regularity lemma proved to be a useful tool. The regularity lemma allows one to import probabilistic intuition to deterministic problems. Moreover, there are many applications where the original problem did not suggest a probabilistic approach.

Motivated especially by questions from computer science, several other variants of Szemerédi's regularity lemma were considered. In Section 2

*The first author was supported by NSF grant DMS 0800070.

we focus mainly on the lemmas proved by Frieze and Kannan [12] and by Alon, Fischer, Krivelevich, and M. Szegedy [2]. We show how these lemmas compare to Szemerédi's original lemma and how they relate to some other variants. Most proofs stated here appeared earlier in the literature and here we just give an overview. A thorough discussion of the connections of those regularity lemmas, from an analytical and geometrical perspective was given recently by Lovász and B. Szegedy in [30]. In Section 3 we discuss the so-called *counting lemmas* and the *removal lemma* and its generalizations. We close with a brief discussion of the *limit approach* of Lovász and B. Szegedy and its relation to the regularity lemmas from Section 2.

There are several surveys devoted to Szemerédi regularity lemma and its applications. The reader is recommended to consult Komlós and Simonovits [26] and Komlós, Shoukoufandeh, Simonovits, and Szemerédi [25], where many applications of the regularity lemma are discussed.

Another line of research, which we will not discuss here, concerns sparse versions of the regularity lemma. Since Szemerédi's lemma is mainly suited for addressing problems involving "dense" graphs, that is graphs with at least $\Omega(|V|^2)$ edges, it is natural to ask for similar statements that would apply to "sparse graphs", i.e., graphs with $o(|V|^2)$ edges. It turns out that a regularity lemma applicable to certain classes of sparse graphs can be proved [22, 34] (see also [1]). Such a lemma was first applied by Kohayakawa and his collaborators to address extremal and Ramsey-type problems for subgraphs of random graphs (see, e.g., [19, 20, 21]). Here we will not further discuss this line of research and we refer the interested reader to the surveys [15, 23, 31] and the references therein.

2. Regularity Lemmas

In this section we discuss several regularity lemmas for graphs. We start our discussion with the regularity lemma of Frieze and Kannan [12] in the next section. In Section 2.2 we show how Szemerédi's regularity lemma [48] can be deduced from the weaker lemma of Frieze and Kannan by iterated applications. In Section 2.3 we discuss the (ε, r)-regularity lemma, whose analog for 3-uniform hypergraphs was introduced by Frankl and Rödl [11]. We continue in Section 2.4 with the regularity lemma of Alon, Fischer, Krivelevich, and M. Szegedy [2], which can be viewed as an iterated version of Szemerédi's regularity lemma. In Section 2.5 we introduce the regular

approximation lemma whose hypergraph variant was developed in [37]. Finally, in Section 2.6 we briefly discuss the original regularity lemma of Szemerédi [47] for bipartite graphs and a multipartite version of it from [8].

2.1. The regularity lemma of Frieze and Kannan

The following variant of Szemerédi's regularity lemma was introduced by Frieze and Kannan [12] for the design of an efficient approximation algorithm for the MAX-CUT problem in dense graphs.

Theorem 1. *For every $\varepsilon > 0$ and every $t_0 \in \mathbb{N}$ there exist $T_{\mathrm{FK}} = T_{\mathrm{FK}}(\varepsilon, t_0)$ and n_0 such that for every graph $G = (V, E)$ with at least $|V| = n \geq n_0$ vertices the following holds. There exists a partition $V_1 \dot\cup \ldots \dot\cup V_t = V$ such that*

(i) $t_0 \leq t \leq T_{\mathrm{FK}}$,

(ii) $|V_1| \leq \cdots \leq |V_t| \leq |V_1| + 1$, *and*

(iii) *for every* $U \subseteq V$

$$(1) \qquad \left| e(U) - \sum_{i=1}^{t-1} \sum_{j=i+1}^{t} d(V_i, V_j) |U \cap V_i| \, |U \cap V_j| \right| \leq \varepsilon n^2,$$

where $e(U)$ denotes the number of edges contained in U and $d(V_i, V_j) = e(V_i, V_j)/(|V_i| \, |V_j|)$ denotes the density of the bipartite graph induced on V_i and V_j.

Definition 2. A partition satisfying property *(ii)* of Theorem 1 will be called *equitable* and a partition satisfying all three properties *(i)*–*(iii)* will be referred to as $(\varepsilon, t_0, T_{\mathrm{FK}})$-*FK-partition*. Sometimes we may omit t_0 and T_{FK} and simply refer to such a partition as ε-*FK-partition*.

The essential properties of the partition provided by Theorem 1 are properties *(i)* and *(iii)*. Property *(i)* bounds the number of partition classes by a constant independent of G and n and, roughly speaking, property *(iii)* asserts that the number of edges of any large set U can be fairly well approximated by the densities $d(V_i, V_j)$ given by the partition $V_1 \dot\cup \ldots \dot\cup V_t = V$. More precisely, $e(U) \approx e(U')$ for any choice of U and U' satisfying for example $|U \cap V_i| \approx |U' \cap V_i|$ for all $i \in [t]$. Moreover, we note that conclusion *(iii)* can be replaced by the following:

(*iii'*) for all (not necessarily disjoint) sets $U, W \subseteq V$

$$\left| e(U,W) - \sum_{i=1}^{t} \sum_{j \in [t] \setminus \{i\}} d(V_i, V_j) |U \cap V_i| \, |W \cap V_j| \right| \leq 6\varepsilon n^2, \tag{2}$$

where edges contained in $U \cap W$ are counted twice in $e(U,W)$.

Indeed, if (*iii*) holds, then we infer (*iii'*) from the identity

$$e(U,W) = e(U \cup W) - e(U) - e(W) + 3e(U \cap W).$$

The proof of Theorem 1 presented here relies on the *index* of a partition, a concept which was first introduced and used by Szemerédi.

Definition 3. For a partition $\mathcal{P} = (V_1, \ldots, V_t)$ of the vertex sets of a graph $G = (V,E)$, i.e., $V_1 \dot\cup \ldots \dot\cup V_t = V$ we define the *index* of $\mathcal{P}$ by

$$\operatorname{ind}(\mathcal{P}) = \frac{1}{\binom{|V|}{2}} \sum_{i=1}^{t-1} \sum_{j=i+1}^{t} d^2(V_i, V_j) |V_i| \, |V_j|.$$

Note that it follows directly from the definition of the index that for any partition $\mathcal{P}$ we have

$$0 \leq \operatorname{ind}(\mathcal{P}) \leq 1.$$

For the proof of Theorem 1 we will use the following consequence of the Cauchy–Schwarz inequality.

Lemma 4. *Let $1 \leq M < N$, let $\sigma_1, \ldots, \sigma_N$ be positive and $d_1, \ldots, d_N$, and d be reals. If $\sum_{i=1}^{N} \sigma_i = 1$ and $d = \sum_{i=1}^{N} d_i \sigma_i$ then*

$$\sum_{i=1}^{N} d_i^2 \sigma_i \geq d^2 + \left(d - \frac{\sum_{i=1}^{M} d_i \sigma_i}{\sum_{i=1}^{M} \sigma_i} \right)^2 \frac{\sum_{i=1}^{M} \sigma_i}{1 - \sum_{i=1}^{M} \sigma_i}.$$

For completeness we include the short proof of Lemma 4.

Proof. For $M = 1$ and $N = 2$ the statement follows from the identity

$$\hat{d}_1^2 \hat{\sigma}_1 + \hat{d}_2^2 \hat{\sigma}_2 = \hat{d}^2 + \left(\hat{d} - \hat{d}_1 \right)^2 \frac{\hat{\sigma}_1}{\hat{\sigma}_2}. \tag{3}$$

which is valid for positive $\hat{\sigma}_1, \hat{\sigma}_2$ with $\hat{\sigma}_1 + \hat{\sigma}_2 = 1$ and $\hat{d} = \hat{d}_1 \hat{\sigma}_1 + \hat{d}_2 \hat{\sigma}_2$.

For general $1 \leq M < N$ we infer from the Cauchy–Schwarz inequality applied twice in the form $\left(\sum d_i\sigma_i\right)^2 \leq \sum d_i^2\sigma_i \sum \sigma_i$

$$\begin{aligned}\sum_{i=1}^{N} d_i^2\sigma_i &= \sum_{i=1}^{M} d_i^2\sigma_i + \sum_{i=M+1}^{N} d_i^2\sigma_i \\ &\geq \frac{\left(\sum_{i=1}^{M} d_i\sigma_i\right)^2}{\sum_{i=1}^{M}\sigma_i} + \frac{\left(\sum_{i=M+1}^{N} d_i\sigma_i\right)^2}{\sum_{i=M+1}^{N}\sigma_i} \\ &= \left(\frac{\sum_{i=1}^{M} d_i\sigma_i}{\sum_{i=1}^{M}\sigma_i}\right)^2 \sum_{i=1}^{M}\sigma_i + \left(\frac{\sum_{i=M+1}^{N} d_i\sigma_i}{\sum_{i=M+1}^{N}\sigma_i}\right)^2 \sum_{i=M+1}^{N}\sigma_i.\end{aligned}$$

Setting

$$\hat{\sigma}_1 = \sum_{i=1}^{M}\sigma_i, \quad \hat{\sigma}_2 = \sum_{i=M+1}^{N}\sigma_i,$$

$$\hat{d}_1 = \frac{\sum_{i=1}^{M} d_i\sigma_i}{\sum_{i=1}^{M}\sigma_i}, \quad \hat{d}_2 = \frac{\sum_{i=M+1}^{N} d_i\sigma_i}{\sum_{i=M+1}^{N}\sigma_i}, \quad \text{and} \quad \hat{d} = \hat{d}_1\hat{\sigma}_1 + \hat{d}_2\hat{\sigma}_2$$

we have $\hat{d} = \sum_{i=1}^{N} d_i\sigma_i = d$ and from (3) we infer

$$\sum_{i=1}^{N} d_i^2\sigma_i \geq \left(\sum_{i=1}^{N} d_i\sigma_i\right)^2 + \left(\sum_{i=1}^{N} d_i\sigma_i - \frac{\sum_{i=1}^{M} d_i\sigma_i}{\sum_{i=1}^{M}\sigma_i}\right)^2 \frac{\sum_{i=1}^{M}\sigma_i}{\sum_{i=M+1}^{N}\sigma_i},$$

which is what we claimed. ■

After those preparations we prove Theorem 1.

Proof of Theorem 1. The proof is based on the following idea already present in the original work of Szemerédi. Starting with an arbitrary equitable vertex partition $\mathcal{P}_0$ with t_0 classes, we consider a sequence of partitions $\mathcal{P}_0, \mathcal{P}_1, \ldots$ such that $\mathcal{P}_j$ always satisfies properties (i) and (ii). As soon as $\mathcal{P}_j$ also satisfies (iii) we can stop. On the other hand, if $\mathcal{P}_j$ does not satisfy (iii) we will show that there exists a partition $\mathcal{P}_{j+1}$ whose index increased by $\varepsilon^2/2$. Since $\operatorname{ind}(\mathcal{P}) \leq 1$ for any partition $\mathcal{P}$, we infer that after at most $2/\varepsilon^2$ steps this procedure must end with a partition satisfying properties (i), (ii), and (iii) of the theorem.

So suppose $\mathcal{P}_j = \mathcal{P} = (V_1, \dots, V_t)$ is a partition of V which satisfies (i) and (ii), but there exists a set $U \subseteq V$ such that (1) fails. We are going to construct a partition $\mathcal{R} = \mathcal{P}_{j+1}$ satisfying

$$\operatorname{ind}(\mathcal{R}) \geq \operatorname{ind}(\mathcal{P}) + \varepsilon^2/2. \tag{4}$$

For that set

$$U_i = V_i \cap U \quad \text{and} \quad \bar{U}_i = V_i \setminus U.$$

We define a new partition $\mathcal{Q}$ by replacing every vertex class V_i by U_i and $\bar{U}_i$

$$\mathcal{Q} = (U_1, \bar{U}_1 \dots, U_t, \bar{U}_t).$$

Next we show that the index of $\mathcal{Q}$ increased by ε^2 compared to $\operatorname{ind}(\mathcal{P})$. For every $1 \leq i < j \leq t$ we set

$$\varepsilon_{ij} = d(U_i, U_j) - d(V_i, V_j).$$

Since we may assume $t \geq t_0 \geq 1/\varepsilon$, which yields $\sum_{i=1}^{t} e(V_i) \leq \varepsilon n^2/2$, we infer from the assumption that (1) fails, that

$$\left| \sum_{i<j} \varepsilon_{ij} |U_i|\,|U_j| \right| \geq \varepsilon n^2 - \sum_{i=1}^{t} e(U_i) \geq \varepsilon n^2 - \sum_{i=1}^{t} e(V_i) \geq \frac{\varepsilon}{2} n^2, \tag{5}$$

Since $V_i = U_i \dot{\cup} \bar{U}_i$ for every $i \in [t]$ we obtain

$$\begin{aligned} d(V_i, V_j)|V_i|\,|V_j| = {} & d(U_i, U_j)|U_i|\,|U_j| + d(\bar{U}_i, U_j)|\bar{U}_i|\,|U_j| \\ & + d(U_i, \bar{U}_j)|U_i|\,|\bar{U}_j| + d(\bar{U}_i, \bar{U}_j)|\bar{U}_i|\,|\bar{U}_j| \end{aligned}$$

and

$$|V_i|\,|V_j| = |U_i|\,|U_j| + |\bar{U}_i|\,|U_j| + |U_i|\,|\bar{U}_j| + |\bar{U}_i|\,|\bar{U}_j|.$$

Combining those identities with Lemma 4, we obtain

$$\begin{aligned} & d^2(U_i, U_j)|U_i|\,|U_j| + d^2(\bar{U}_i, U_j)|\bar{U}_i|\,|U_j| + d^2(U_i, \bar{U}_j)|U_i|\,|\bar{U}_j| \\ & + d^2(\bar{U}_i, \bar{U}_j)|\bar{U}_i|\,|\bar{U}_j| \geq d^2(V_i, V_j)|V_i|\,|V_j| + \varepsilon_{ij}^2 \left(\frac{|U_i|\,|U_j|}{1 - \frac{|U_i|\,|U_j|}{|V_i|\,|V_j|}} \right) \\ & \geq d^2(V_i, V_j)|V_i|\,|V_j| + \varepsilon_{ij}^2 |U_i|\,|U_j|. \end{aligned}$$

Summing over all $1 \leq i < j \leq t$ we obtain

$$\text{(6)} \quad \operatorname{ind}(\mathcal{Q}) \geq \operatorname{ind}(\mathcal{P}) + \frac{1}{\binom{n}{2}} \sum_{i<j} \varepsilon_{ij}^2 |U_i|\,|U_j|$$

$$\geq \operatorname{ind}(\mathcal{P}) + \frac{\left(\sum_{i<j} \varepsilon_{ij}|U_i|\,|U_j|\right)^2}{\binom{n}{2}\sum_{i<j}|U_i|\,|U_j|} \overset{(5)}{\geq} \operatorname{ind}(\mathcal{P}) + \frac{(\varepsilon n^2/2)^2}{\binom{n}{2}\binom{n}{2}} \geq \operatorname{ind}(\mathcal{P}) + \varepsilon^2.$$

We now find an equitable partition $\mathcal{R}$ which is a refinement of $\mathcal{P}$ (and almost a refinement of $\mathcal{Q}$) for which (4) holds. For that subdivide each vertex class V_i of $\mathcal{P}$ into sets $W_{i,a}$ of size $\lfloor \varepsilon^2 n/(5t) \rfloor$ or $\lfloor \varepsilon^2 n/(5t) \rfloor + 1$ in such a way that for all but at most one of these sets either $W_{i,a} \subseteq U_i$ or $W_{i,a} \subseteq \bar{U}_i$ holds. For every $i \in [t]$ let $W_{i,0}$ denote the exceptional set if it exists and let $W_{i,0}$ be arbitrary otherwise. Let $\mathcal{R}$ be the resulting partition. Moreover, we consider the partition $\mathcal{R}^*$ which is a refinement of $\mathcal{R}$ obtained by replacing $W_{i,0}$ by possibly two classes $U_i \cap W_{i,0}$ and $\bar{U}_i \cap W_{i,0}$. Since the contribution of the index of $\mathcal{R}$ and $\mathcal{R}^*$ may differ only on pairs with at least one vertex in $W_{i,0}$ for some $i \in [t]$ and since $|W_{i,0}| \leq \lfloor \varepsilon^2 n/(5t) \rfloor + 1$ for every $i \in [t]$ we infer that

$$\operatorname{ind}(\mathcal{R}^*) - \operatorname{ind}(\mathcal{R}) \leq \binom{n}{2}^{-1} \sum_{i=1}^{t} \left(\frac{\varepsilon^2 n}{5t} + 1 \right) n \leq \frac{\varepsilon^2}{2}.$$

for sufficiently large n. Furthermore, since $\mathcal{R}^*$ is a refinement of $\mathcal{Q}$ it follows from the Cauchy–Schwarz inequality that $\operatorname{ind}(\mathcal{Q}) \leq \operatorname{ind}(\mathcal{R}^*)$ and, consequently,

$$\operatorname{ind}(\mathcal{R}) \geq \operatorname{ind}(\mathcal{R}^*) - \frac{\varepsilon^2}{2} \geq \operatorname{ind}(\mathcal{Q}) - \frac{\varepsilon^2}{2} \overset{(6)}{\geq} \operatorname{ind}(\mathcal{P}) + \frac{\varepsilon^2}{2},$$

which concludes the proof of the theorem. ■

The proof of Theorem 1 shows that choosing

$$T_{\mathrm{FK}}(\varepsilon, t_0) = \max\{t_0, 1/\varepsilon\} \cdot (6/\varepsilon^2)^{2/\varepsilon^2} = t_0 2^{\mathrm{poly}(1/\varepsilon)}$$

suffices. In fact, in each refinement step we split the vertex classes V_i into at most $\lfloor 5/\varepsilon^2 + 1 \rfloor \leq 6/\varepsilon^2$ classes $W_{i,a}$, when we construct $\mathcal{R}$. Hence, each time property *(iii)* fails the number of vertex classes of the new partition increases by a factor of $6/\varepsilon^2$ and in total there are at most $2/\varepsilon^2$ iterations.

On the other hand, it was shown by Lovász and B. Szegedy [30] that for every $0 < \varepsilon \leq 1/3$ there are graphs for which every partition into t classes satisfying property (iii) of Theorem 1 requires $t \geq 2^{1/(8\varepsilon)}/4$ and, hence, $t \gg 1/\varepsilon$. As a consequence Theorem 1 does not allow to obtain useful bounds for $e(U \cap V_i, U \cap V_j)$, since for such a graph $\varepsilon n^2 \gg n^2/t^2 = |V_i|\,|V_j|$. Property (iii) of Theorem 1 only implies $e(U \cap V_i, U \cap V_j) \approx d(V_i, V_j)|U \cap V_i|\,|U \cap V_j|$ on average over all pairs $i < j$ for every "large" set U. However, Szemerédi's regularity lemma (which was proved long before Theorem 1) allows to control $e(U \cap V_i, U \cap V_j)$ for most $i < j$. The price of this is, however, a significantly larger upper bound for the number of partition classes t.

2.2. Szemerédi's regularity lemma

In this section we show how Szemerédi's regularity lemma from [48] can be obtained from Theorem 1 by iterated applications. For that we consider the following simple corollary of Theorem 1, which was first considered by Tao [49].

Corollary 5. *For all ν, $\varepsilon > 0$, every function $\delta : \mathbb{N} \to (0,1]$, and every $t_0 \in \mathbb{N}$ there exist $T_0 = T_0\big(\nu, \varepsilon, \delta(\cdot), t_0\big)$ and n_0 such that for every graph $G = (V, E)$ with at least $|V| = n \geq n_0$ vertices the following holds. There exists a vertex partition $\mathcal{P} = (V_i)_{i\in[t]}$ with $V_1 \dot\cup \ldots \dot\cup V_t = V$ and a refinement $\mathcal{Q} = (W_{i,j})_{i\in[t],\, j\in[s]}$ with $W_{i,1} \dot\cup \ldots \dot\cup W_{i,s} = V_i$ for every $i \in [t]$ such that*

- (i) *$\mathcal{P}$ is an (ε, t_0, T_0)-FK-partition,*
- (ii) *$\mathcal{Q}$ is a $\big(\delta(t), t_0, T_0\big)$-FK-partition, and*
- (iii) *$\operatorname{ind}(\mathcal{Q}) \leq \operatorname{ind}(\mathcal{P}) + \nu$.*

Before we deduce Corollary 5 from Theorem 1, we discuss property (iii). Roughly speaking, if two refining partitions $\mathcal{P}$ and $\mathcal{Q}$ satisfy property (iii), then this implies that $d(W_{i,a}, W_{j,b})$ and $d(V_i, V_j)$ are "relatively close" for "most" choices of $i < j$ and $a, b \in [s]$. More precisely, we have the following, which was already observed by Alon, Fischer, Krivelevich, and M. Szegedy [2].

Lemma 6. *Let γ, $\nu > 0$, let $G = (V, E)$ be a graph with n vertices, and for some positive integers t and s let $\mathcal{P} = (V_i)_{i\in[t]}$ with $V_1 \dot\cup \ldots \dot\cup V_t = V$*

be a vertex partition and let $\mathcal{Q} = (W_{i,j})_{i\in[t],\ j\in[s]}$ *be a refinement with* $W_{i,1}\dot\cup\ldots\dot\cup W_{i,s} = V_i$ *for every* $i \in [t]$. *If* $\operatorname{ind}(\mathcal{Q}) \le \operatorname{ind}(\mathcal{P}) + \nu$, *then*

$$\sum_{1\le i<j\le t}\ \sum_{a,b\in[s]} \Big\{|W_{i,a}|\,|W_{j,b}| \,:\, \big|d(W_{i,a},W_{j,b}) - d(V_i,V_j)\big| \ge \gamma\Big\} \le \frac{\nu}{\gamma^2}n^2.$$

Proof. For $1 \le i < j \le t$ let $A_{ij}^+ = \big\{(a,b) \in [s]\times[s] \,:\, d(W_{i,a},W_{j,b}) - d(V_i,V_j) \ge \gamma\big\}$. Since

$$d(V_i,V_j)|V_i|\,|V_j| = \sum_{a,b\in[s]} d(W_{i,a},W_{j,b})|W_{i,a}|\,|W_{j,b}|$$

$$= \sum_{(a,b)\in A_{ij}^+} d(W_{i,a},W_{j,b})|W_{i,a}|\,|W_{j,b}| + \sum_{(a,b)\notin A_{ij}^+} d(W_{i,a},W_{j,b})|W_{i,a}|\,|W_{j,b}|,$$

we obtain from the defect form of Cauchy–Schwarz (Lemma 4), that

$$\sum_{a,b\in[s]} d^2(W_{i,a},W_{j,b})|W_{i,a}|\,|W_{j,b}| \ge d^2(V_i,V_j)|V_i|\,|V_j| + \gamma^2 \sum_{(a,b)\in A_{ij}^+} |W_{i,a}|\,|W_{j,b}|.$$

Summing over all $1 \le i < j \le t$ we get

$$\operatorname{ind}(\mathcal{Q}) \ge \operatorname{ind}(\mathcal{P}) + \frac{\gamma^2}{\binom{n}{2}} \sum_{1\le i<j\le t}\ \sum_{(a,b)\in A_{ij}^+} |W_{i,a}|\,|W_{j,b}|.$$

Since, by assumption $\operatorname{ind}(\mathcal{Q}) \le \operatorname{ind}(\mathcal{P}) + \nu$, we have

$$\sum_{1\le i<j\le t}\ \sum_{(a,b)\in A_{ij}^+} |W_{i,a}|\,|W_{j,b}| \le \frac{\nu}{\gamma^2}\binom{n}{2} \le \frac{\nu n^2}{2\gamma^2}.$$

Repeating the argument with the appropriate definition of A_{ij}^- yields the claim. ∎

Proof of Corollary 5. For the proof of the corollary we simply iterate Theorem 1. Without loss of generality we may assume that $\delta(t) \le \varepsilon$ for every $t \in \mathbb{N}$. For given ν, ε, $\delta(\cdot)$, and t_0, we apply Theorem 1 and obtain an (ε, t_0, T_0)-FK-partition $\mathcal{P}$ with t classes. Since in the proof of Theorem 1 the initial partition was an arbitrary equitable partition, we infer that after

another application of Theorem 1 with $\delta(t)$ (in place of ε) and t_0 we obtain an equitable refinement $\mathcal{Q}$ of $\mathcal{P}$ which is a $(\delta(t), t_0, T_0)$-FK-partition with st classes. In other words, $\mathcal{P}$ and $\mathcal{Q}$ satisfy properties (i) and (ii) of Corollary 5 and if (iii) also holds, then we are done. On the other hand, if (iii) fails, then we replace $\mathcal{P}$ by $\mathcal{Q}$ and iterate, i.e., we apply Theorem 1 with $\delta(ts)$ (in place of ε) and $t_0 = ts$ to obtain an equitable refinement $\mathcal{Q}'$ of $\mathcal{P}' = \mathcal{Q}$. Since we only iterate as long as (iii) of Corollary 5 fails and since ν is fixed throughout the proof, this procedure must end after at most $1/\nu$ iterations. Therefore the upper bound T_0 on the number of classes is in fact independent of G and n and can be given by a recursive formula depending on ν, ε, $\delta(\cdot)$, and t_0. ■

We now show that Corollary 5 applied with the right choice of parameters yields the following theorem, which is essentially Szemerédi's regularity lemma from [48].

Theorem 7. *For every $\varepsilon > 0$ and every $t_0 \in \mathbb{N}$ there exist $T_{\mathrm{Sz}} = T_{\mathrm{Sz}}(\varepsilon, t_0)$ and n_0 such that for every graph $G = (V, E)$ with at least $|V| = n \geq n_0$ vertices the following holds. There exists a partition $V_1 \dot{\cup} \ldots \dot{\cup} V_t = V$ such that*

- (i) $t_0 \leq t \leq T_{\mathrm{Sz}}$,
- (ii) $|V_1| \leq \cdots \leq |V_t| \leq |V_1| + 1$, *and*
- (iii) *for all but at most εt^2 pairs (V_i, V_j) with $i < j$ we have that for all subsets $U_i \subseteq V_i$ and $U_j \subseteq V_j$*

$$\big|e(U_i, U_j) - d(V_i, V_j)|U_i|\,|U_j|\big| \leq \varepsilon |V_i|\,|V_j|. \tag{7}$$

We note that the usual statement of Szemerédi's regularity lemma is slightly different from the one above. Usually $\varepsilon|V_i|\,|V_j|$ on the right-hand side of (7) is replaced by $\varepsilon|U_i|\,|U_j|$ and for (iii) it is assumed that $|U_i| \geq \varepsilon|V_i|$ and $|U_j| \geq \varepsilon|V_j|$. However, applying Theorem 7 with $\varepsilon' = \varepsilon^3$ would yield a partition with comparable regular properties.

Definition 8. Pairs (V_i, V_j) for which (7) holds for every $U_i \subseteq V_i$ and $U_j \subseteq V_j$ are called *ε-regular*. Partitions satisfying all three properties (i)–(iii) of Theorem 7, we will refer to as *$(\varepsilon, t_0, T_{\mathrm{Sz}})$-Szemerédi-partition*. Again we may sometimes omit t_0 and T_{Sz} and simply refer to such partitions as *ε-Szemerédi-partitions*.

Below we deduce Theorem 7 from Corollary 5 and Lemma 6.

Proof of Theorem 7. For given $\varepsilon > 0$ and t_0, we apply Corollary 5 with

$$\nu' = \frac{\varepsilon^4}{36^2}, \quad \varepsilon' = 1, \quad \delta'(t) = \frac{\varepsilon}{36t^2}, \quad \text{and} \quad t_0' = t_0$$

and obtain constants T_0' and n_0' which define $T_{\mathrm{Sz}} = T_0'$ and $n_0 = n_0'$. (We remark that the choice for ε' has no bearing for the proof and therefore we set it equal to 1.) For a given graph $G = (V, E)$ with n vertices Corollary 5 yields two partitions $\mathcal{P} = (V_i)_{i\in[t]}$ and $\mathcal{Q} = (W_{i,j})_{i\in[t],\ j\in[s]}$ satisfying properties (i)–(iii) of Corollary 5. We will show that, in fact, the coarser partition $\mathcal{P}$ also satisfies properties (i)–(iii) of Theorem 7. Since $\mathcal{P}$ is an $(\varepsilon', t_0', T_0')$-FK-partition by our choice of $t_0' = t_0$ and $T_{\mathrm{Sz}} = T_0'$ the partition $\mathcal{P}$ obviously satisfies properties (i) and (ii) of Theorem 7 and we only have to verify property (iii).

For that we consider for every $1 \le i < j \le t$ the set

$$A_{ij} = \Big\{(a,b) \in [s] \times [s] \,:\, \big|d(W_{i,a}, W_{j,b}) - d(V_i, V_j)\big| \ge \varepsilon/6\Big\}$$

and we let

$$I = \left\{\{i,j\} \,:\, 1 \le i < j \le t \text{ such that } \sum_{(a,b)\in A_{ij}} |W_{i,a}|\,|W_{j,b}| \ge \varepsilon |V_i|\,|V_j|/6\right\}.$$

We will first show that $|I| \le \varepsilon t^2$ and then we will verify that if $\{i,j\} \notin I$, then (7) holds. Indeed, due to property (iii) of Corollary 5 we have $\mathrm{ind}\,(\mathcal{Q}) \le \mathrm{ind}\,(\mathcal{P}) + \nu'$ and, consequently, it follows from Len ma 6 (applied with $\nu' = \varepsilon^4/36^2$ and $\gamma' = \varepsilon/6$) that

$$\begin{aligned}\frac{\varepsilon^2 n^2}{36} &\ge \sum_{i<j} \sum_{(a,b)\in A_{ij}} |W_{i,a}|\,|W_{j,b}| \\ &\ge \sum_{\{i,j\}\in I} \sum_{(a,b)\in A_{ij}} |W_{i,a}|\,|W_{j,b}| \ge \frac{\varepsilon}{6} \sum_{\{i,j\}\in I} |V_i|\,|V_j|.\end{aligned}$$

Moreover, since $|V_i| \ge \lfloor n/t \rfloor \ge n/(2t)$ for every $i \in [t]$ we have $\varepsilon n^2/6 \ge |I| n^2/(4t^2)$ and, consequently,

$$|I| \le \frac{2}{3}\varepsilon t^2 < \varepsilon t^2. \tag{8}$$

Next we will show that if $\{i,j\} \notin I$ then the pair (V_i, V_j) is ε-regular, i.e., we show that (7) holds for every $U_i \subseteq V_i$ and $U_j \subseteq V_j$. For given sets $U_i \subseteq V_i$ and $U_j \subseteq V_j$ and $a, b \in [s]$ we set

$$U_{i,a} = U_i \cap W_{i,a} \quad \text{and} \quad U_{j,b} = U_j \cap W_{j,b}$$

and have

$$e(U_i, U_j) = \sum_{a,b\in[s]} e(U_{i,a}, U_{j,b}).$$

Appealing to the fact that $\mathcal{Q}$ is a $\left(\delta'(t), t_0', T_0'\right)$-FK-partition we obtain from (2) that

$$e(U_i, U_j) = \sum_{a,b\in[s]} d(W_{i,a}, W_{j,b})|U_{i,a}|\,|U_{j,b}| \pm 6\delta'(t)n^2.$$

From the assumption $\{i,j\} \notin I$ we infer

$$\sum_{(a,b)\in A_{ij}} d(W_{i,a}, W_{j,b})|U_{i,a}|\,|U_{j,b}| \le \sum_{(a,b)\in A_{ij}} |W_{i,a}|\,|W_{j,b}| \le \frac{\varepsilon}{6}|V_i|\,|V_j|$$

and, furthermore, for $(a,b) \notin A_{ij}$ we have

$$d(W_{i,a}, W_{j,b})|U_{i,a}|\,|U_{j,b}| = \left(d(V_i, V_j) \pm \frac{\varepsilon}{6}\right)|U_{i,a}|\,|U_{j,b}|.$$

Combining, those three estimates we infer

$$e(U_i, U_j) = \sum_{a,b\in[s]} d(V_i, V_j)|U_{i,a}|\,|U_{j,b}| \pm \frac{\varepsilon}{6}|U_i|\,|U_j| \pm \frac{\varepsilon}{6}|V_i|\,|V_j| \pm 6\delta'(t)n^2.$$

Hence from our choice of $\delta'(t)$ and $V_i \ge \lfloor n/t \rfloor \ge n/(2t)$ we deduce

$$\left|e(U_i, U_j) - d(V_i, V_j)|U_i|\,|U_j|\right| \le \frac{\varepsilon}{3}|V_i|\,|V_j| + \frac{\varepsilon}{6}\left(\frac{n}{t}\right)^2 \le \varepsilon|V_i|\,|V_j|,$$

which concludes the proof of Theorem 7. ∎

In contrast to Theorem 1 the upper bound $T_{\mathrm{Sz}} = T_{\mathrm{Sz}}(\varepsilon, t_0)$ we obtain from the proof of Theorem 7 is not exponential, but of tower-type. In fact, we use Corollary 5 with $\nu = \varepsilon^4/36^2$ and $\delta(t) = \varepsilon/(36t^2)$. Due to the choice of ν we iterate Theorem 1 at most $36^2/\varepsilon^4$ times and each time the number of classes grows exponentially, i.e., t_i classes from the i-th iteration may split

into $2^{O(t_i^4/\varepsilon^2)}$ classes for the next step. As a consequence, the upper bound $T_{\mathrm{Sz}} = T_{\mathrm{Sz}}(\varepsilon, t_0)$, which we obtain from this proof, is a tower of 4's of height $O(\varepsilon^{-4})$ with t_0 as the last exponent. The proof of Szemerédi's regularity lemma from [48] yields a similar upper bound of a tower of 2's of height proportional to ε^{-5}. However, recall that the statement from [48] is slightly different from the version proved here, by having a smaller error term in (7). A lower bound of similar type was obtained by Gowers [17]. In fact, Gowers showed an example of a graph for which any partition satisfying even only a considerably weaker version of property (*iii*) requires at least t classes, where t is a tower of 2's of height proportional to $1/\varepsilon^{1/16}$.

2.3. The (ε, r)-regularity lemma

As we have just discussed in the previous section, the example of Gowers shows that we cannot prevent the situation when the number of parts t of a Szemerédi-partition is much larger than, say, $1/\varepsilon$. For several applications this presents an obstacle which one would like to overcome. More precisely one would like to obtain some control of the densities of subgraphs which are of size much smaller than, say, n/t^2. The (ε, r)-regularity lemma (Theorem 9), the regularity lemma of Alon, Fischer, Krivelevich, and M. Szegedy (Theorem 10), and the regular approximation lemma (Theorem 11), were partly developed to address such issues.

A version for 3-uniform hypergraphs of the following regularity lemma was obtained by Frankl and Rödl in [11].

Theorem 9. *For every $\varepsilon > 0$, every function $r : \mathbb{N} \to \mathbb{N}$, and every $t_0 \in \mathbb{N}$ there exist $T_{\mathrm{FR}} = T_{\mathrm{FR}}(\varepsilon, r(\cdot), t_0)$ and n_0 such that for every graph $G = (V, E)$ with at least $|V| = n \geq n_0$ vertices the following holds. There exists a partition $V_1 \dot\cup \ldots \dot\cup V_t = V$ such that*

(*i*) $t_0 \leq t \leq T_{\mathrm{FR}}$,

(*ii*) $|V_1| \leq \cdots \leq |V_t| \leq |V_1| + 1$, *and*

(*iii*) *for all but at most εt^2 pairs (V_i, V_j) with $i < j$ we have that for all sequences of subsets $U_i^1, \ldots, U_i^{r(t)} \subseteq V_i$ and $U_j^1, \ldots, U_j^{r(t)} \subseteq V_j$*

$$\left| \left| \bigcup_{q=1}^{r(t)} E(U_i^q, U_j^q) \right| - d(V_i, V_j) \left| \bigcup_{q=1}^{r(t)} U_i^q \times U_j^q \right| \right| \leq \varepsilon |V_i|\,|V_j|. \tag{9}$$

Note that if $r(t) \equiv 1$ then Theorem 9 is identical to Theorem 7 and if $r(t) \equiv k$ for some constant $k \in \mathbb{N}$ (independent of t), then it is a direct consequence of Theorem 7. We remark that for arbitrary functions $r(\cdot)$, Theorem 9 can be proved along the lines of Szemerédi's proof of Theorem 7 from [48]. Below we deduce Theorem 9, using a slightly different approach, namely we infer Theorem 9 from Corollary 5 in a similar way as we proved Theorem 7.

Proof. For given ε, $r(\cdot)$, and t_0 we follow the lines of the proof of Theorem 7. This time we apply Corollary 5 with a smaller choice of $\delta'(\cdot)$

$$\nu' = \frac{\varepsilon^4}{36^2}, \quad \varepsilon' = 1, \quad \delta'(t) = \frac{\varepsilon}{36t^2\big(4^{r(t)} - 3^{r(t)}\big)}, \quad \text{and} \quad t_0' = t_0$$

and obtain T_0' and n_0', which determines T_{FR} and n_0. We define the sets A_{ij} and I identical as in the proof of Theorem 7, i.e., for $1 \le i < j \le t$ we set

$$A_{ij} = \Big\{(a,b) \in [s] \times [s] \,:\, \big|d(W_{i,a}, W_{j,b}) - d(V_i, V_j)\big| \ge \varepsilon/6\Big\}$$

and we let

$$I = \left\{\{i,j\} \,:\, 1 \le i < j \le t \text{ such that } \sum_{(a,b)\in A_{ij}} |W_{i,a}|\,|W_{j,b}| \ge \varepsilon |V_i|\,|V_j|/6\right\}.$$

Again we obtain (8) and the rest of the proof requires some small straightforward adjustments.

We set $r = r(t)$ and we will show that if $\{i,j\} \notin I$, then (9) holds for every sequence $\hat{U}_i^1, \dots, \hat{U}_i^r \subseteq V_i$ and $\hat{U}_j^1, \dots, \hat{U}_j^r \subseteq V_j$. For such given sequences we consider new sequences $U_i^1, \dots, U_i^R \subseteq V_i$ and $U_j^1, \dots, U_j^R \subseteq V_j$ satisfying the disjointness property (see (10) below). For that let $R = 4^r - 3^r$ and for a non-empty set $\emptyset \neq L \subseteq [r]$ let

$$\hat{U}_i(L) = \bigcap_{\ell\in L} \hat{U}_i^\ell \setminus \bigcup_{\ell\in L} \hat{U}_i^\ell \quad \text{and} \quad \hat{U}_j(L) = \bigcap_{\ell\in L} \hat{U}_j^\ell \setminus \bigcup_{\ell\in L} \hat{U}_j^\ell$$

and for two sets L, L' with non-empty intersection we let $U_i(L,L') = \hat{U}_i(L)$ and $U_j(L,L') = \hat{U}_j(L')$. Note that there are $R = 4^r - 3^r$ such pairs of sets L, L' and we can relabel the sequences $\big(U_i(L,L')\big)_{L\cap L'\neq\emptyset}$ and $\big(U_j(L,L')\big)_{L\cap L'\neq\emptyset}$ to $U_i^1, \dots, U_i^R \subseteq V_i$ and $U_j^1, \dots, U_j^R \subseteq V_j$. Note that for

all $p \neq q$ the sets U_i^p and U_i^q may either be equal or disjoint. Moreover, due to this definition we obtain for all $1 \leq p < q \leq R$

$$(10)\quad \left(U_i^q \times U_j^q\right) \cap \left(U_i^p \times U_j^p\right) = \emptyset \quad \text{and} \quad \dot{\bigcup}_{q\in[R]} U_i^q \times U_j^q = \bigcup_{q\in[r]} \hat{U}_i^q \times \hat{U}_j^q.$$

Furthermore, for $q \in [R]$ and $a, b \in [s]$ we set

$$U_{i,a}^q = U_i^q \cap W_{i,a} \quad \text{and} \quad U_{j,b}^q = U_j^q \cap W_{j,b}$$

and we get for every $q \in [R]$

$$e\left(U_i^q, U_j^q\right) = \sum_{a,b\in[s]} e\left(U_{i,a}^q, U_{j,b}^q\right).$$

Appealing to the fact that $\mathcal{Q}$ is a $\left(\delta'(t), t_0', T_0'\right)$-FK-partition we obtain from (2) that

$$e\left(U_i^q, U_j^q\right) = \sum_{a,b\in[s]} d\left(W_{i,a}, W_{j,b}\right) \left|U_{i,a}^q\right| \left|U_{j,b}^q\right| \pm 6\delta'(t)n^2.$$

From the assumption $\{i,j\} \notin I$ and the disjointness property from (10) we infer

$$\sum_{(a,b)\in A_{ij}} \sum_{q\in[R]} d(W_{i,a}, W_{j,b}) \left|U_{i,a}^q\right| \left|U_{j,b}^q\right| \leq \sum_{(a,b)\in A_{ij}} |W_{i,a}|\,|W_{j,b}| \leq \frac{\varepsilon}{6}|V_i|\,|V_j|$$

and, furthermore, for $(a,b) \notin A_{ij}$ we have

$$d(W_{i,a}, W_{j,b}) \left|U_{i,a}^q\right| \left|U_{j,b}^q\right| = \left(d(V_i, V_j) \pm \frac{\varepsilon}{6}\right) \left|U_{i,a}^q\right| \left|U^{rq}_{j,b}\right|$$

for every $q \in [R]$. Combining, those three estimates we infer

$$\begin{aligned}
&\left|\bigcup_{q\in[r]} E\left(\hat{U}_i^q, \hat{U}_j^q\right)\right| = \left|\dot{\bigcup}_{q\in[R]} E\left(U_i^q, U_j^q\right)\right| \\
&= \left(d(V_i, V_j) \pm \frac{\varepsilon}{6}\right) \sum_{q\in[R]} \sum_{a,b\in[s]} \left|U_{i,a}^q\right| \left|U_{j,b}^q\right| \pm \frac{\varepsilon}{6}|V_i|\,|V_j| \pm 6R\delta'(t)n^2 \\
&= d(V_i, V_j)\left|\bigcup_{q\in[R]} E\left(U_i^q, U_j^q\right)\right| \pm \frac{\varepsilon}{3}|V_i|\,|V_j| \pm 6R\delta'(t)n^2.
\end{aligned}$$

Hence from our choice of $\delta'(t)$, $R = (4^r - 3^r)$, and $V_i \geq \lfloor n/t \rfloor \geq n/(2t)$ we deduce from (10)

$$\left| \left| \bigcup_{q=1}^{r} E(\hat{U}_i^q, \hat{U}_j^q) \right| - d(V_i, V_j) \left| \bigcup_{q=1}^{r} \hat{U}_i^q \times \hat{U}_j^q \right| \right| \leq \varepsilon |V_i|\,|V_j|,$$

which concludes the proof of Theorem 9. ■

2.4. The regularity lemma of Alon, Fischer, Krivelevich, and M. Szegedy

In the last two sections we iterated the regularity lemma of Frieze and Kannan and obtained Corollary 5, from which we deduced Szemerédi's regularity lemma (Theorem 7) and the (ε, r)-regularity lemma (Theorem 9).

From this point of view it seems natural to iterate these stronger regularity lemmas. This was indeed first carried out by Alon, Fischer, Krivelevich, and M. Szegedy [2] who iterated Szemerédi's regularity lemma for an application in the area of property testing.

Theorem 10. *For every ν, $\varepsilon > 0$, every function $\delta : \mathbb{N} \to (0, 1]$, and every $t_0 \in \mathbb{N}$ there exist $T_{\mathrm{AFKS}} = T_{\mathrm{AFKS}}\big(\nu, \varepsilon, \delta(\cdot), t_0\big)$ and n_0 such that for every graph $G = (V, E)$ with at least $|V| = n \geq n_0$ vertices the following holds. There exists a vertex partition $\mathcal{P} = (V_i)_{i \in [t]}$ with $V_1 \dot{\cup} \ldots \dot{\cup} V_t = V$ and a refinement $\mathcal{Q} = (W_{i,j})_{i \in [t],\, j \in [s]}$ with $W_{i,1} \dot{\cup} \ldots \dot{\cup} W_{i,s} = V_i$ for every $i \in [t]$ such that*

(i) $\mathcal{P}$ is an $(\varepsilon, t_0, T_{\mathrm{AFKS}})$–Szemerédi-partition,

(ii) $\mathcal{Q}$ is a $\big(\delta(t), t_0, T_{\mathrm{AFKS}}\big)$–Szemerédi-partition, and

(iii) $\mathrm{ind}\,(\mathcal{Q}) \leq \mathrm{ind}\,(\mathcal{P}) + \nu$.

Proof. The proof is identical to the proof of Corollary 5 with the only adjustment that we iterate Theorem 7 instead of Theorem 1. ■

The price for the stronger properties of the partitions $\mathcal{P}$ and $\mathcal{Q}$, in comparison to Szemerédi's regularity lemma, is again in the bound T_{AFKS}. In general T_{AFKS} can be expressed as a recursive formula in $\nu, \varepsilon, \delta(\cdot)$, and

t_0, and for example, if $\delta(t)$ is given by a polynomial in $1/t$, then T_{AFKS} is an iterated tower-type function, which is sometimes referred to as a wowzer-type function.

Theorem 9 relates to Theorem 10 in the following way. It is a direct consequence of (9) that if (V_i, V_j) is not one of the exceptional pairs in (iii) of Theorem 9, then for any partition of V_i and V_j into at most $\sqrt{r(t)}$ parts of equal size, "most" of the $r(t)$ pairs have the density "close" (up to an error of $O = \big(\sqrt{\varepsilon}\,\big)$) to $d(V_i, V_j)$. Hence, if we set at the beginning $r(t) = \big(T_{\mathrm{Sz}}\big(\delta(t), t\big)\big)^2$ and then apply Theorem 7 to obtain a $\big(\delta(t), t, T_{\mathrm{Sz}}\big(\delta(t), t\big)\big)$–Szemerédi-partition $\mathcal{Q}$, which refines the given partition, then we arrive to a similar situation as in Theorem 10. In fact, we have two Szemerédi-partitions satisfying (i) and (ii) of Theorem 10 and (iii) would be replaced by the fact that $d(W_{i,a}, W_{j,b}) \approx d(V_i, V_j)$ for "most" pairs from the finer partition $\mathcal{Q}$.

2.5. The regular approximation lemma

The following regularity lemma is another byproduct of the hypergraph generalization of the regularity lemma and appeared in general form in [37]. In a different context, Theorem 11 appeared in the work of Lovász and B. Szegedy [30, Lemma 5.2].

Theorem 11. *For every $\nu > 0$, every function $\varepsilon : \mathbb{N} \to (0, 1]$, and every $t_0 \in \mathbb{N}$ there exist $T_0 = T_0(\nu, \varepsilon(\cdot), t_0)$ and n_0 such that for every graph $G = (V, E)$ with at least $|V| = n \geq n_0$ vertices the following holds. There exists a partition $\mathcal{P} = (V_i)_{i \in [t]}$ with $V_1 \dot\cup \ldots \dot\cup V_t = V$ and a graph $H = (V, E')$ on the same vertex set V as G such that*

(a) *$\mathcal{P}$ is an $\big(\varepsilon(t), t_0, T_0\big)$–Szemerédi-partition for H and*

(b) *$|E \triangle E'| = |E \setminus E'| + |E' \setminus E| \leq \nu n^2$.*

The main difference between Theorem 11 and Theorem 7 is in the choice of ε being a function of t. As already mentioned, it follows from the work of Gowers [17] (or alternatively from the work of Lovász and B. Szegedy [30, Proposition 7.1]) that it is not possible to obtain a Szemerédi (or even a Frieze-Kannan) partition for certain graphs G with ε of order $1/t$. Property (a) of Theorem 11 asserts, however, that by adding and deleting at most νn^2 edges from/to G we can obtain another graph H which admits a "much more" regular partition, e.g., with $\varepsilon(t) \ll 1/t$.

Below we show how Theorem 11 can be deduced from the iterated regularity lemma of Alon, Fischer, Krivelevich and M. Szegedy (Theorem 10). The idea is to apply Theorem 10 with appropriate parameters to obtain Szemerédi-partitions $\mathcal{P} = (V_i)_{i\in[t]}$ and $\mathcal{Q} = (W_{i,j})_{i\in[t],\ j\in[s]}$ for which $\mathcal{Q}$ refines $\mathcal{P}$ and $\operatorname{ind}(\mathcal{Q}) \leq \operatorname{ind}(\mathcal{P}) + \nu'$. The last condition and Lemma 6 imply that $d(W_{i,a}, W_{j,b}) \approx d(V_i, V_j)$ (with an error depending on ν') for "most" $i < j$ and $a, b \in [s]$. The strong regularity of the finer partition $\mathcal{Q}$ will then be used to adjust G (by adding and removing a few edges randomly) to obtain H for which $\mathcal{P}$ will have the desired properties. We now give the details of this outline.

Proof of Theorem 11. For given ν, $\varepsilon(\cdot)$, and t_0 we apply Theorem 10 with $\nu' = \nu^3/16$, some arbitrary ε', say $\varepsilon' = 1$, $\delta'(t) = \min\{\varepsilon(t)/2, \nu/4\}$, and $t_0' = t_0$. We also fix an auxiliary constant $\gamma' = \nu/2$. We then set $T_0 = T'_{\mathrm{AFKS}}$ and $n_0 = n_0'$. After we apply Theorem 10 to the given graph $G = (V, E)$, we obtain an (ε', t_0, T_0)–Szemerédi-partition $\mathcal{P}$ and a $(\delta'(t), t_0, T_0)$–Szemerédi-partition $\mathcal{Q}$ which refines $\mathcal{P}$ such that $\operatorname{ind}(\mathcal{Q}) \leq \operatorname{ind}(\mathcal{P}) + \nu'$.

Next we will change G and obtain the graph H, which will satisfy (a) and (b) of Theorem 11. For that:

(A) we replace every subgraph $G[W_{i,a}, W_{j,b}]$ which is not $\delta'(t)$-regular by a random bipartite graph of density $d(V_i, V_j)$ and

(B) for every $1 \leq i < j \leq t$ and $a, b \in [s]$ we add or remove edges randomly to change the density of $G[W_{i,a}, W_{j,b}]$ to $d(V_i, V_j) + o(1)$.

It follows from the Chernoff bound that the resulting graph $H = (V, E')$ has the property that for *every* $1 \leq i < j \leq t$ and $a, b \in [s]$ the induced subgraph $H[W_{i,a}, W_{j,b}]$ is $\big(\delta'(t)+o(1)\big)$-regular and $d_H(W_{i,a}, W_{j,b}) = d_H(V_i, V_j)+o(1)$, where $o(1) \to 0$ as $n \to \infty$. (Recall that for G from Lemma 6 we can only infer that $d_G(W_{i,a}, W_{j,b}) = d_G(V_i, V_j) \pm \gamma'$ for "most" pairs for some $\gamma' \gg \delta'(t)$.) Hence for every $1 \leq i < j \leq t$ and arbitrary sets $U_i \subseteq V_i$ and $U_j \subseteq V_j$ we have

$$
\begin{aligned}
&e_H(U_i, U_j) \\
&= \sum_{a,b\in[s]} \big(d_H(W_{i,a}, W_{j,b})|U_i \cap W_{i,a}|\,|U_j \cap W_{j,b}| \pm \big(\delta'(t) + o(1)\big)\,|W_{i,a}|\,|W_{j,b}|\big) \\
&= d_H(V_i, V_j)|U_i|\,|U_j| \pm 2\delta'(t)|V_i|\,|V_j|.
\end{aligned}
$$

In other words, the partition $\mathcal{P}$ is a $\big(2\delta'(t) \leq \varepsilon(t), t_0, T_0\big)$–Szemerédi-partition for H, which is assertion (a) of Theorem 11. For part (b) we will estimate the symmetric difference of E and E'. Since $\mathcal{Q}$ is a $\big(\delta'(t), t_0, T_0\big)$–Szemerédi-partition for G the changes in Step (A) contributed at most

$$\delta'(t)t^2s^2\left\lceil\frac{n}{ts}\right\rceil^2 \leq \frac{\nu}{2}n^2 \tag{11}$$

to that difference.

For estimating the changes introduced in Step (B) we appeal to Lemma 6.

From that we infer that, since $\operatorname{ind}(\mathcal{Q}) \leq \operatorname{ind}(\mathcal{P}) + \nu'$, we "typically" changed only

$\gamma'|W_{i,a}||W_{j,b}|$ pairs. More precisely, in Step (B) we changed at most

$$\begin{aligned}(12)\qquad &\sum_{i<j}\sum_{a,b\in[s]}\gamma'|W_{i,a}|\,|W_{j,b}| \\ &+\sum_{i<j}\sum_{a,b\in[s]}\big\{|W_{i,a}|\,|W_{j,b}| : \big|d_G(W_{i,a},W_{j,b}) - d_G(V_i,V_j)\big| \geq \gamma'\big\} \\ &\leq \left(\frac{\gamma'}{2} + \frac{\nu'}{(\gamma')^2}\right)n^2 \leq \frac{\nu}{2}n^2.\end{aligned}$$

Finally, from (11) and (12) we infer $|E\triangle E'| \leq \nu n^2$, which shows that H satisfies property (b) of Theorem 11. ■

2.6. An early version of the regularity lemma

We close this section with the statement of an early version of Szemerédi's regularity lemma, which was introduced in [47] and one of the key components in the proof of the Erdős-Turán conjecture concerning the upper density of subsets of the integers containing no arithmetic progression of fixed length. Another application of that lemma lead to the upper bound for the Ramsey-Turán problem for K_4 due to Szemerédi [46] and to the resolution of the $(6,3)$-problem, which was raised by Brown, Erdős and Sós [6, 45], and solved by Ruzsa and Szemerédi [43].

Theorem 12. *For all positive ε_1, ε_2, δ, ϱ, and σ there exist T_0, S_0, M, and N such that for every bipartite graph $G = (X \dot\cup Y, E)$ satisfying $|X| = m \geq M$ and $|Y| = n \geq N$ there exists a partition $X_0 \dot\cup X_1 \dot\cup \ldots \dot\cup X_t = X$ with $t \leq T_0$ and for every $i = 1, \ldots, t$ there exists a partition $Y_{i,0} \dot\cup Y_{i,1} \dot\cup \ldots \dot\cup Y_{i,s_i} = Y$ with $s_i \leq S_0$ such that*

(a) *$|X_0| \leq \varrho m$ and $|Y_{i,0}| \leq \sigma n$ for every $i = 1, \ldots, t$, and*

(b) *for every $i = 1, \ldots, t$, every $j = 1, \ldots, s_i$, and all sets $U \subseteq X_i$ and $W \subseteq Y_{i,j}$ with $|U| \geq \varepsilon_1 |X_i|$ and $|W| \geq \varepsilon_2 |Y_{i,j}|$ we have $d(U, W) \geq d(X_i, Y_{i,j}) - \delta$.* ■

Note that this lemma does not ensure such an easy to use structure of the partition as the later lemmas. More precisely, the partitions of Y may be very different for every $i = 1, \ldots, t$. On the other hand, the upper bounds T_0 and S_0 are of similar type as those of Theorem 1, i.e., we have $T_0, S_0 = 2^{\text{poly}\left(1/\min\{\varepsilon_1, \varepsilon_2, \delta, \varrho, \sigma\}\right)}$. We also point out that for example in [43] Theorem 12 was applied iteratively, which in turn lead to a tower-type bound for the $(6,3)$-problem. A multipartite version of Theorem 12 was developed by Duke, Lefmann, and Rödl [8] for efficiently approximating the subgraph frequencies in a given graph G on n vertices for subgraphs of up to $\Omega\big(\sqrt{\log \log (n)}\big)$ vertices.

Theorem 13. *For every $\varepsilon > 0$ and every integer $k \geq 2$ there exist $T_0 = 4^{k^2/\varepsilon^5}$ such that for every k-partite graph $G = (V, E)$ with vertex classes $V_1 \dot\cup \ldots \dot\cup V_k = V$ and $|V_1| = \cdots = |V_k| = N$ there exists a partition $\mathcal{P}$ of $V_1 \times \cdots \times V_k$ such that*

(i) *the number of elements $W_1 \times \cdots \times W_k$ in $\mathcal{P}$ is at most T_0,*

(ii) *$|W_i| \geq \varepsilon^{k^2/\varepsilon^5} N$ for every $i = 1, \ldots, k$ and every $W_1 \times \cdots \times W_k$ in $\mathcal{P}$, and*

(iii) *we have*

$$\sum_{W_1 \times \cdots \times W_k \in \mathcal{P}_{\text{irr}}} \prod_{i=1}^{k} |W_i| \leq \varepsilon N^k.$$

for the subfamily $\mathcal{P}_{\text{irr}} \subseteq \mathcal{P}$ containing those elements $W_1 \times \cdots \times W_k$ from $\mathcal{P}$ which contain an irregular pair (W_i, W_j), i.e., a pair (W_i, W_j) with $i < j$ for which there exist subsets $U_i \subseteq W_i$ and $U_j \subseteq W_j$ with $|U_i| \geq \varepsilon |W_i|$ and $|U_j| \geq \varepsilon |W_j|$ such that $\big|d(U_i, U_j) - d(W_i, W_j)\big| > \varepsilon$. ■

The main advantage of Theorem 13, in comparison to Szemerédi's regularity lemma (Theorem 7), is the smaller upper bound T_0. The partition in Theorem 13 still conveys information if $1/\varepsilon$ and k tend slowly to infinity with $n = |V|$, for example, if $1/\varepsilon$ and k are of order $\log^c(n)$ for some small constant $c > 0$. Due to the tower-type bound of Theorem 7 there $1/\varepsilon$ can be at most of order $\log^*(n)$, where $\log^*$ denotes the iterated logarithm function.

On the other hand, the upper bound T_0 in Theorem 13 is comparable to the one from Theorem 1 and as we will see in the next section Theorem 1 would be also well suited for the main application of Theorem 13 in [8]. Moreover, the structure of the partition provided by Theorem 1 seems to be simpler and easier to work with.

3. Reduced Graph and Counting Lemmas

In this section we show how regular properties of the partitions given by the regularity lemmas from Section 2 can be applied to approximate the number of subgraphs of fixed isomorphism type of a given graph G. More precisely, for graphs G and F let $N_F(G)$ denote the number of labeled copies of F in G. Roughly speaking, we will show that $N_F(G)$ can be fairly well approximated by only studying the so-called *reduced graph* (or *cluster-graph*) of a regular partition.

Definition 14. Let $\varepsilon > 0$, $G = (V, E)$ be a graph, and let $\mathcal{P} = (V_i)_{i\in[t]}$ be a partition of V.

(*i*) If $\mathcal{P}$ is an ε-FK-partition, then the *reduced graph* $R = R_G(\mathcal{P})$ is defined to be the weighted, complete, undirected graph with vertex set $V(R) = [t]$ and edge weights $w_R(i,j) = d(V_i, V_j)$.

(*ii*) If $\mathcal{P}$ is an ε–Szemerédi-partition, then the *reduced graph* $R = R_G(\mathcal{P}, \varepsilon)$ is defined to be the weighted, undirected graph with vertex set $V(R) = [t]$, edge set $E(R) = \big\{\{i,j\} : (V_i, V_j) \text{ is } \varepsilon\text{-regular}\big\}$, and edge weights $w_R(i,j) = d(V_i, V_j)$.

The reduced graph carries a lot of the structural information of the given graph G. In fact, in many applications of the regularity lemma, the original problem for G one is interested in can be turned into a "simpler" problem for the reduced graph.

Remark 15. Below we will consider (labeled) copies F_R of a given graph F in a reduced graph R. If R is the reduced graph of an FK-partition, then R is an edge-weighting of the complete graph and, consequently, any ordered set of $\big|V(F)\big|$ vertices of $V(R)$ spans a copy of F. On the other hand, if R is the reduced graph of an ε–Szemerédi-partition, then R is not a complete graph and for a labeled copy F_R of F in R with $V(F_R) = \{i_1, \dots, i_\ell\}$ we will have that (V_{i_j}, V_{i_k}) is ε-regular for every edge $\{i_j, i_k\} \in E(F_R)$.

3.1. The global counting lemma

Here by a *counting lemma* we mean an assertion which enables us to deduce directly from the reduced graph some useful information on the number $N_F(G)$ of labeled copies of a fixed graph F in a large graph G. We will distinguish between two different settings here. The first counting lemma will yield an estimate on $N_F(G)$ in the context of Theorem 1. Since $N_F(G)$ concerns the total number of copies, we regard this result as a *global counting lemma.*

In contrast, for an ℓ-vertex graph F the *local counting lemma* (Theorem 18) will yield estimates on $N_F\big(G[V_{i_1}, \dots, V_{i_\ell}]\big)$ for an induced ℓ-partite subgraph of G given by the regular partition $\mathcal{P}$. However, for such a stronger assertion we will require that $\mathcal{P}$ be a Szemerédi-partition.

Theorem 16. *Let F be a graph with vertex set $V(F) = [\ell]$. For every $\gamma > 0$ there exists $\varepsilon > 0$ such that for every $G = (V, E)$ with $|V| = n$ and every ε-FK-partition $\mathcal{P} = (V_i)_{i \in [t]}$ with reduced graph $R = R_G(\mathcal{P})$ we have*

$$N_F(G) = \sum_{F_R} \prod_{\{i_j, i_k\} \in E(F_R)} w_R(i_j, i_k) \prod_{i_j \in V(F_R)} |V_{i_j}| \pm \gamma n^\ell, \tag{13}$$

where the sum runs over all labeled copies F_R of F in R (cf. Remark 15).

For a simpler notation we denote here and below the vertices $V(F_R)$ of a given copy of F_R of the ℓ-vertex graph F in R by $\{i_1, \dots, i_\ell\}$ and omit the dependence of F_R.

Proof. We follow an argument of Lovász and B. Szegedy from [29]. We prove Theorem 18 by induction on the number of edges of F. Clearly, the theorem holds for graphs with no edges and for graphs with one edge it follows from the definition of ε-FK-partition with $\varepsilon = \gamma$.

For given F and γ we let $\varepsilon \le \gamma/12$ be sufficiently small, so that the statement for the induction assumption holds with $\gamma' = \gamma/2$. For two vertices $x, y \in V$ we set

$$d_{\mathcal{P}}(x,y) = \begin{cases} 0 & \text{if } x, y \in V_i \text{ for some } i \in [t], \\ d(V_i, V_j) & \text{if } x \in V_i \text{ and } y \in V_j \text{ for some } 1 \le i < j \le t \end{cases}$$

and we denote by $\mathbb{1}_E(x,y)$ the indicator function for E, i.e., $\mathbb{1}_E(x,y)$ equals 1 if $\{x,y\} \in E$ and it equals 0 otherwise. We consider the difference of the left-hand side and the main term of the right-hand side in (13) and obtain

$$\left| N_F(G) - \sum_{F_R} \prod_{\{i_j,i_k\} \in E(F_R)} w_R(i_j,i_k) \prod_{i_j \in V(F_R)} |V_{i_j}| \right| \tag{14}$$

$$= \left| \sum_{x_1,\dots,x_\ell \in (V)_\ell} \left(\prod_{\{i,j\} \in E(F)} \mathbb{1}_E(x_i,x_j) - \prod_{\{i,j\} \in E(F)} d_{\mathcal{P}}(x_i,x_j) \right) \right|,$$

where $x_1, \dots, x_\ell \in (V)_\ell$ is an arbitrary sequence of ℓ distinct vertices in V. Without loss of generality we may assume that $\{\ell-1, \ell\}$ is an edge in F and we denote by F^- the spanning subgraph of F with the edge $\{\ell-1,\ell\}$ removed. Then, applying the identity $\alpha_1\alpha_2 - \beta_1\beta_2 = \beta_2(\alpha_1 - \beta_1) + \alpha_1(\alpha_2 - \beta_2)$, we get the following upper bound for the right-hand side of the last equation

$$\left| \sum_{x_1,\dots,x_\ell \in (V)_\ell} d_{\mathcal{P}}(x_{\ell-1},x_\ell) \left(\prod_{\{i,j\} \in E(F^-)} \mathbb{1}_E(x_i,x_j) - \prod_{\{i,j\} \in E(F^-)} d_{\mathcal{P}}(x_i,x_j) \right) \right| \tag{15}$$

$$+ \left| \sum_{x_1,\dots,x_\ell \in (V)_\ell} \left(\prod_{\{i,j\} \in E(F^-)} \mathbb{1}_E(x_i,x_j) \right) \left(\mathbb{1}_E(x_{\ell-1},x_\ell) - d_{\mathcal{P}}(x_{\ell-1},x_\ell) \right) \right|.$$

By the induction assumption we can bound the first term by $\gamma' n^\ell$, i.e., we have

$$\left| \sum_{x_1,\dots,x_\ell \in (V)_\ell} d_{\mathcal{P}}(x_{\ell-1},x_\ell) \left(\prod_{\{i,j\} \in E(F^-)} \mathbb{1}_E(x_i,x_j) - \prod_{\{i,j\} \in E(F^-)} d_{\mathcal{P}}(x_i,x_j) \right) \right| \tag{16}$$

$$\le \gamma' n^\ell.$$

We will verify a similar bound for the second term in (15). For that we will split the second term of (15) into two parts and rewrite each of the parts (see (17) and (18) below).

We consider the induced subgraph F^* of F, which we obtain by removing the vertices labeled $\ell-1$ and ℓ from F. For a copy $\tilde{F}^*$ of F^* in G let $X_{\ell-1}(\tilde{F}^*)$ and $X_\ell(\tilde{F}^*)$ be those vertex sets such that for every pair $x_{\ell-1} \in X_{\ell-1}(\tilde{F}^*)$ and $x_\ell \in X_\ell(\tilde{F}^*)$ of distinct vertices, those two vertices extend $\tilde{F}^*$ in G to a copy of F^-. More precisely, if $x_1, \dots, x_{\ell-2}$ is the vertex set of $\tilde{F}^*$ then we set

$$X_{\ell-1}(\tilde{F}^*) = \bigcap_{i:\{i,\ell-1\}\in E(F)} \Gamma_G(x_i) \quad \text{and} \quad X_\ell(\tilde{F}^*) = \bigcap_{i:\{i,\ell\}\in E(F)} \Gamma_G(x_i),$$

where $\Gamma_G(x)$ denotes the set of neighbours of x in G. To simplify the notation, below we will write $X_{\ell-1}$ or X_ℓ instead of $X_{\ell-1}(\tilde{F}^*)$ or $X_\ell(\tilde{F}^*)$ as $\tilde{F}^*$ will be clear from the context. Since by definition edges contained in $X_{\ell-1} \cap X_\ell$ are counted twice in $e(X_{\ell-1}, X_\ell)$ (cf. (2)) we observe for the first part of the second term in (15) that

$$\sum_{x_1,\dots,x_\ell\in(V)_\ell} \Bigg(\prod_{\{i,j\}\in E(F^-)} \mathbb{1}_E(x_i,x_j) \Bigg) \mathbb{1}_E(x_{\ell-1},x_\ell) = \sum_{\tilde{F}^*} e(X_{\ell-1},X_\ell), \tag{17}$$

Moreover, we have for the second part of the seond term in (15)

$$\begin{aligned} \sum_{x_1,\dots,x_\ell\in(V)_\ell} \Bigg(\prod_{\{i,j\}\in E(F^-)} \mathbb{1}_E(x_i,x_j) \Bigg) d_{\mathcal{P}}(x_{\ell-1},x_\ell) \\ = \sum_{\tilde{F}^*} \sum_{i\neq j\in[t]} d(V_i,V_j)|X_{\ell-1}\cap V_i|\,|X_\ell\cap V_j| \end{aligned} \tag{18}$$

and, consequently, we can bound the second term in (15) by

$$\sum_{\tilde{F}^*} \Bigg| e(X_{\ell-1},X_\ell) - \sum_{i\neq j\in[t]} d(V_i,V_j)|X_{\ell-1}\cap V_i|\,|X_\ell\cap V_j| \Bigg| \tag{19}$$

Finally, we can apply the fact that $\mathcal{P}$ is an ε-FK-partition in form of (2) and the fact that $N_{F^*}(G) \le n^{\ell-2}$ to bound (19) by $n^{\ell-2}\cdot 6\varepsilon n^2$. Hence, from (14)–(19) we infer

$$\Bigg| N_F(G) - \sum_{F_R} \prod_{\{i_j,i_k\}\in E(F_R)} w_R(i_j,i_k) \prod_{i_j\in V(F_R)} |V_{i_j}| \Bigg| \le (\gamma' + 6\varepsilon)n^\ell \le \gamma n^\ell,$$

which concludes the proof of Theorem 18. ∎

A simple argument based on the principle of inclusion and exclusion yields an induced version of Theorem 16. Let $N_F^*(G)$ denote the number of labeled, induced copies of F in G.

Corollary 17. *Let F be a graph with vertex set $V(F) = [\ell]$. For every $\gamma > 0$ there exists $\varepsilon > 0$ such that for every $G = (V, E)$ with $|V| = n$ and every ε-FK-partition $\mathcal{P} = (V_i)_{i \in [t]}$ with reduced graph $R = R_G(\mathcal{P})$ we have*

$$N_F^*(G) = \sum_{F_R} \prod_{\{i_j, i_k\} \in E(F_R)} w_R(i_j, i_k) \prod_{\{i_j, i_k\} \in E(\overline{F_R})} \left(1 - w_R(i_j, i_k)\right) \times \prod_{i_j \in V(F_R)} |V_{i_j}| \pm \gamma n^\ell,$$

where the sum runs over all labeled copies F_R of F in R and $\overline{F_R}$ denotes the complement graph of F_R on the same ℓ vertices $V(F_R)$.

Proof. Let F be a graph with $V(F) = [\ell]$ and let K^ℓ be the complete graph on the same vertex set. Let ε be sufficiently small, so that we can apply Theorem 16 with $\gamma' = \gamma / 2^{\binom{\ell}{2} - e(F)}$ for every graph $F' \subseteq K^\ell$ which contains F. Let G, an ε-FK-partition $\mathcal{P}$, and a reduced graph $R = R_G(\mathcal{P})$ be given.

Due to the principle of inclusion and exclusion we have

$$N_F^*(G) = \sum_{F \subseteq F' \subseteq K^\ell} (-1)^{e(F') - e(F)} N_{F'}(G),$$

where we sum over all supergraphs F' of F contained in K^ℓ. Applying Theorem 16 for every such F' we obtain

$$N_F^*(G) = \sum_{F'} (-1)^{e(F') - e(F)} \left(\sum_{F'_R} \prod_{\{i_j, i_k\} \in E(F'_R)} w_R(i_j, i_k) \prod_{i_j \in V(F_R)} |V_{i_j}| \right) \pm \gamma n^\ell,$$

where the outer sum runs over all F' with $F \subseteq F' \subseteq K^\ell$ and the inner sum is indexed by all copies F'_R of F' in R. We can rewrite the main term by rearranging the sum in the following way: First we sum over all possible labeled copies F_R of F in R. Note that this fixes a unique labeled copy

$K^\ell(F_R)$ of K^ℓ as well, and in the inner sum we consider all graphs F'_R in R "sandwiched" between F_R and $K^\ell(F_R)$. This way we obtain

$$\begin{aligned}
&N^*_F(G) \pm \gamma n^\ell \\
&= \sum_{F_R} \sum_{F_R \subseteq F'_R \subseteq K^\ell(F_R)} (-1)^{e(F'_R)-e(F_R)} \prod_{\{i_j,i_k\}\in E(F'_R)} w_R(i_j,i_k) \prod_{i_j \in V(F_R)} |V_{i_j}| \\
&= \sum_{F_R} \prod_{\{i_j,i_k\}\in E(F_R)} w_R(i_j,i_k) \prod_{i_j\in V(F_R)} |V_{i_j}| \\
&\quad \times \sum_{F_R \subseteq F'_R \subseteq K^\ell(F_R)} (-1)^{e(F'_R)-e(F_R)} \prod_{\{i_j,i_k\}\in E(F'_R)\setminus E(F_R)} w_R(i_j,i_k) \\
&= \sum_{F_R} \prod_{\{i_j,i_k\}\in E(F_R)} w_R(i_j,i_k) \prod_{i_j\in V(F_R)} |V_{i_j}| \\
&\quad \times \prod_{\{i_j,i_k\}\in E\left(K^\ell(F_R)\right)\setminus E(F_R)} \big(1 - w_R(i_j,i_k)\big),
\end{aligned}$$

which concludes the proof. ■

3.2. The local counting lemma

For graphs F and G, a partition $\mathcal{P} = (V_i)_{i\in[t]}$ of $V(G)$, and a labeled copy F_R of F in R with $V(F_R) = \{i_1, \dots, i_\ell\}$ we denote by $N_F\big(G[F_R]\big)$ the number of *partite isomorphic-copies* of F_R (and hence of F) in G induced on $V_{i_1} \dot\cup \dots \dot\cup V_{i_\ell}$. In other words, $N_F\big(G[F_R]\big)$ is the number of edge preserving mappings φ from $V(F_R)$ to $V_{i_1} \dot\cup \dots \dot\cup V_{i_\ell}$ such that $\varphi(i_j) \in V_{i_j}$ for every $j = 1, \dots, \ell$.

Roughly speaking, the global counting lemma from the last section asserts that if $\mathcal{P}$ is a sufficiently regular ε-FK-partition, then $N_G(F)$ can be estimated from the reduced graph $R_G(\mathcal{P})$. In fact, it follows that the average of $N_F\big(G[F_R]\big)$ over all labeled copies F_R of F in R is "close" to its expectation. The local counting lemma (Theorem 18), states that if $\mathcal{P}$ is, in fact, a sufficiently regular Szemerédi-partition, then this is not only true on average, but indeed for every copy F_R of F in R.

Recall, that by definition the edge set $E(R)$ of a reduced graph of a Szemerédi-partition $\mathcal{P}$ corresponds to the regular pairs of $\mathcal{P}$. Consequently, for a copy F_R of F in R we require that all edges of F_R correspond to regular pairs.

Theorem 18. *Let F be a graph with ℓ vertices. For every $\gamma > 0$ there exists $\varepsilon > 0$ such that for every $G = (V, E)$ with $|V| = n$ and every ε–Szemerédi-partition $\mathcal{P} = (V_i)_{i\in[t]}$ with reduced graph $R = R_G(\mathcal{P}, \varepsilon)$ we have for every labeled copy F_R of F in R with $V(F_R) = \{i_1, \ldots, i_\ell\}$*

$$N_F\big(G[F_R]\big) = \prod_{\{i_j,i_k\}\in E(F_R)} w_R(i_j, i_k) \prod_{i_j\in V(F_R)} |V_{i_j}| \pm \gamma \prod_{i_j\in V(F_R)} |V_{i_j}|.$$

Theorem 18 concerns the number of copies of a fixed graph F and will only give interesting bounds if we can assert $\gamma \ll \prod_{\{i_j,i_k\}\in E(F_R)} w_R(i_j, i_k)$. Moreover, it was shown by Chvátal, Rödl, Szemerédi, and Trotter [7], that if H is a graph of bounded degree with cn/t vertices (for some appropriate $c > 0$ depending on $\min_{\{i_j,i_k\}\in E(F_R)} w_R(i_j, i_k)$ and $\Delta(H)$) and there exists a homomorphism from H into F_R, then, under the same assumptions as in Theorem 18, $G[F_R]$ contains a copy of H. A far reaching strengthening, the so-called *blow-up lemma*, was found by Komlós, Sárközy, and Szemerédi [24]. The blow-up lemma allows, under some slightly more restrictive assumptions, to embed spanning graphs H of bounded degree.

Proof. We prove Theorem 18 by induction on the number of edges of F. Since the theorem is trivial for graphs with no edges and it follows from the definition of ε–Szemerédi-partition for $\varepsilon = \gamma$ for graphs with precisely one edge.

Let F be a graph with at least two edges and ℓ vertices. For given $\gamma > 0$ let $\varepsilon \le \gamma/2$ be sufficiently small, so that the theorem holds for F^- with $\gamma' = \gamma/2$. Let $G = (V, E)$ be given along with an ε–Szemerédi-partition $\mathcal{P} = (V_i)_{i\in[t]}$ and let F_R be a labeled copy of F in R. Without loss of generality we may assume that $V(F_R) = \{1, \ldots, \ell\}$ and that $\{\ell - 1, \ell\}$ is an edge of F_R. We denote by F_R^- the subgraph of F_R which we obtain after deleting the edge $\{\ell - 1, \ell\}$ from F_R. We can express the number of partite isomorphic copies of F_R through

$$N_F\big(G[F_R]\big) = \sum_{x_1\in V_1} \cdots \sum_{x_\ell\in V_\ell} \prod_{\{i,j\}\in E(F_R)} \mathbb{1}_E(x_i, x_j)$$

$$= \sum_{x_1 \in V_1} \cdots \sum_{x_\ell \in V_\ell} \prod_{\{i,j\} \in E(F_R^-)} \big(\mathbb{1}_E(x_i, x_j)$$

$$\times \big(d(V_{\ell-1}, V_\ell) + \mathbb{1}_E(x_{\ell-1}, x_\ell) - d(V_{\ell-1}, V_\ell) \big) \big).$$

The last expression can be rewritten as

$$d(V_{\ell-1}, V_\ell) \times N_{F^-}\big(G[F_R^-]\big)$$

$$+ \sum_{x_1 \in V_1} \cdots \sum_{x_\ell \in V_\ell} \Bigg(\prod_{\{i,j\} \in E(F_R^-)} \mathbb{1}_E(x_i, x_j) \big(\mathbb{1}_E(x_{\ell-1}, x_\ell) - d(V_{\ell-1}, V_\ell) \big) \Bigg).$$

From the induction assumption we then infer

$$d(V_{\ell-1}, V_\ell) N_F\big(G[F_R^-]\big)$$

$$= \prod_{\{i_j, i_k\} \in E(F_R)} w_R(i_j, i_k) \prod_{i_j \in V(F_R)} |V_{i_j}| \pm \frac{\gamma}{2} \prod_{i_j \in V(F_R)} |V_{i_j}|$$

and, therefore, it suffices to verify

$$\text{(20)} \quad \Bigg| \sum_{x_1 \in V_1} \cdots \sum_{x_\ell \in V_\ell} \Bigg(\prod_{\{i,j\} \in E(F_R^-)} \mathbb{1}_E(x_i, x_j) \big(\mathbb{1}_E(x_{\ell-1}, x_\ell) - d(V_{\ell-1}, V_\ell) \big) \Bigg) \Bigg|$$

$$\leq \frac{\gamma}{2} \prod_{i_j \in V(F_R)} |V_{i_j}|.$$

For that we will appeal to the regularity of $\mathcal{P}$. Let F_R^* be the induced subgraph of F_R which one obtains by removing the vertices $\ell - 1$ and ℓ. For a partite isomorphic copy $\tilde{F}^*$ of F_R^*, let $X_{\ell-1}(\tilde{F}^*) \subseteq V_{\ell-1}$ and $X_\ell(\tilde{F}^*) \subseteq V_\ell$ be those sets of vertices for which any choice of $x_{\ell-1} \in X_{\ell-1}(\tilde{F}^*)$ and $x_\ell \in X_\ell(\tilde{F}^*)$ complete $\tilde{F}^*$ to a partite isomorphic copy of F_R^-. (To simplify the notation, below we will write $X_{\ell-1}$ or X_ℓ instead of $X_{\ell-1}(\tilde{F}^*)$ or $X_\ell(\tilde{F}^*)$ as $\tilde{F}^*$ will be clear from the context.) Consequently, summing over all partite isomorphic copies $\tilde{F}^*$ of F_R^* in G we obtain

$$\Bigg| \sum_{x_1 \in V_1} \cdots \sum_{x_\ell \in V_\ell} \Bigg(\prod_{\{i,j\} \in E(F_R^-)} \mathbb{1}_E(x_i, x_j) \Big(\mathbb{1}_E(x_{\ell-1}, x_\ell) - d(V_{\ell-1}, V_\ell) \Big) \Bigg) \Bigg|$$

$$= \left| \sum_{\tilde{F}^*} e(X_{\ell-1}, X_\ell) - d(V_{\ell-1}, V_\ell)|X_{\ell-1}|\,|X_\ell| \right|$$

$$\leq \sum_{\tilde{F}^*} \left| e(X_{\ell-1}, X_\ell) - d(V_{\ell-1}, V_\ell)|X_{\ell-1}|\,|X_\ell| \right|$$

$$\leq \prod_{i=1}^{\ell-2} |V_i| \times \varepsilon |V_{\ell-1}|\,|V_\ell|,$$

where in the last estimate we used the ε-regularity of $(V_{\ell-1}, V_\ell)$ and the obvious upper bound on the number of partite isomorphic copies $\tilde{F}^*$ of F_R^*. Since $\varepsilon \leq \gamma/2$ the assertion (20) follows and concludes the proof of Theorem 18. ■

We close this section by noting that an induced version of Theorem 18 can be derived directly from Theorem 18 in a similar way as Corollary 17 (we omit the details).

Corollary 19. *Let F be a graph with ℓ vertices. For every $\gamma > 0$ there exists $\varepsilon > 0$ such that for every $G = (V, E)$ with $|V| = n$ and every ε–Szemerédi-partition $\mathcal{P} = (V_i)_{i \in [t]}$ with reduced graph $R = R_G(\mathcal{P}, \varepsilon)$ the following is true.*

For every labeled copy F_R of F contained in a clique $K_R^\ell \subseteq R$ with $V(F_R) = V(K_R^\ell) = \{i_1, \ldots, i_\ell\}$

$$\prod_{\{i_j, i_k\} \in E(F_R)} w_R(i_j, i_k) \prod_{\{i_j, i_k\} \in \binom{V(F_R)}{2} \setminus E(F_R)} \big(1 - w_R(i_j, i_k)\big) \prod_{i_j \in V(F_R)} |V_{i_j}|$$

$$= N_F^*\big(G[F_R]\big) \pm \gamma \prod_{i_j \in V(F_R)} |V_{i_j}|,$$

where $N_F^*\big(G[F_R]\big)$ denotes the number of labeled, induced, partite isomorphic copies of F_R in $G[F_R] = G[V_{i_1} \dot{\cup} \ldots \dot{\cup} V_{i_\ell}]$. ■

Note that by assumption of Corollary 19 and the definition of the reduced graph for Szemerédi-partitions we require for $F_R \subseteq R$ with $V(F_R) = \{i_1, \ldots, i_\ell\}$, that (V_{i_j}, V_{i_k}) is ε-regular for every pair $\{i_j, i_k\}$ and not only for pairs corresponding to edges of F_R.

3.3. The removal lemma and its generalizations

A direct consequence of the local counting lemma is the so-called *removal lemma*. Answering a question of Brown, Sós, and Erdős [6, 45] Ruzsa and Szemerédi [43] established the *triangle removal lemma*. They proved that every graph which contains only $o(n^3)$ triangles can be made triangle free by removing at most $o(n^2)$ edges. This result was generalized by Erdős, Frankl, and Rödl [10] from triangles to arbitrary graphs.

Theorem 20 (Removal lemma for graphs). *For every graph F with ℓ vertices and every $\eta > 0$ there exists $c > 0$ and n_0 such that every graph $G = (V, E)$ on $n \geq n_0$ vertices with $N_F(G) < cn^\ell$, there exists a subgraph $H = (V, E')$ such that $N_F(H) = 0$ and $|E \setminus E'| \leq \eta n^2$.*

While the original proof of Ruzsa and Szemerédi was based on an iterated application of the early version of the regularity lemma, Theorem 12, the proof given in [10] is based on Szemerédi's regularity lemma, Theorem 7. We remark that even in the triangle case both proofs give essentially the same tower-type dependency between c and η, i.e., c is a polynomial in $1/T$, where T is a tower of 2's of height polynomial in $1/\eta$. It is an intriguing open problem to find a proof which gives a better dependency between c and η.

Proof. Suppose that $G = (V, E)$ is a graph which even after the deletion of any set of at most ηn^2 edges still contains a copy of F. We will show that such a graph G contains at least cn^ℓ copies of F. For that we apply Szemerédi's regularity lemma, Theorem 7, with

$$\varepsilon = \min\left\{\frac{\eta}{8\ell^2}, \frac{\varepsilon'}{3\ell^2}\right\} \quad \text{and} \quad t_0 = \frac{1}{\eta},$$

where ε' is given by the local counting lemma applied with F and

$$\gamma = \frac{1}{3}\left(\frac{\eta}{4}\right)^{e(F)},$$

and obtain an ε–Szemerédi-partition $\mathcal{P} = (V_i)_{i\in[t]}$ of V. Next we delete all edges $e \in E$ for which at least one of the following holds:

- $e \subseteq V_i$ for some $i \in [t]$,
- $e \in E(V_i, V_j)$ for some $1 \leq i < j \leq t$ such that (V_i, V_j) is not ε-regular,

- $e \in E(V_i, V_j)$ for some $1 \leq i < j \leq t$ such that $d(V_i, V_j) \leq \eta/2$.

Simple calculations show that we delete at most ηn^2 edges in total. Let G' be the graph, which we obtain after the deletion of those edges. Due to the assumption on G, the graph G' must still contain a copy F_0 of F. Therefore the reduced graph $R = R_{G'}(\mathcal{P}, \varepsilon)$ must contain a copy of a homomorphic image F'_R of F for which $w_R(i_j, j_k) \geq \eta/2$ for all $\{i_j, j_k\} \in E(F'_R)$.

If F'_R is a copy of F, then the local counting lemma, Theorem 18, implies that G' contains, for sufficiently large n at least

$$\left(\frac{\eta}{2}\right)^{e(F)} \left\lfloor \frac{n}{t} \right\rfloor^{\ell} - \gamma \left\lceil \frac{n}{t} \right\rceil^{\ell} \geq \frac{1}{2} \left(\frac{\eta}{2}\right)^{e(F)} \left(\frac{n}{t}\right)^{\ell}$$

copies of F. Consequently, $N_F(G) \geq N_F(G') \geq cn^{\ell}$, for some c only depending on η and $T_{\mathrm{Sz}}\big(\min\{\eta/(8\ell^2), \varepsilon'/(3\ell^2)\}, 1/\eta\big)$, where ε' only depends on F and η. In other words, there exists such a c which only depends on the graph F and η as claimed.

The case when F'_R is not isomorphic to F is very similar. For example, we may subdivide every vertex class V_i into ℓ classes, $V_{i,1} \dot{\cup} \ldots \dot{\cup} V_{i,\ell} = V_i$, and obtain a refinement $\mathcal{Q}$. It follows from the definition of ε-regular pair, that if (V_i, V_j) is ε-regular, then $(V_{i,a}, V_{j,b})$ is $(3\ell^2\varepsilon)$-regular for any $a, b \in [\ell]$ and $d(V_{i,a}, V_{j,b}) \geq d(V_i, V_j) - 2\ell^2\varepsilon$. Since F'_R was contained in R, the reduced graph $S = S_{G'}(\mathcal{Q}, 3\ell^2\varepsilon)$ must contain a full copy F_R of F for which $w_R(i_j, j_k) \geq \eta/2 - 2\ell^2\varepsilon \geq \eta/4$ for all $\{i_j, j_k\} \in E(F_R)$ and the local counting lemma yields $N_F(G) \geq cn^{\ell}$ for

$$c = \frac{1}{2\ell^{\ell}} \left(\frac{\eta}{4}\right)^{e(F)} T_{\mathrm{Sz}}\left(\min\left\{\frac{\eta}{8\ell^2}, \frac{\varepsilon'}{3\ell^2}\right\}, \frac{1}{\eta}\right)^{-\ell}. \qquad \blacksquare$$

It was shown by Ruzsa and Szemerédi [43] that the removal lemma for triangles can be used to deduce Szemerédi's theorem on arithmetic progressions for progressions of length 3, which was earlier (and with better quantitative bounds) proved by Roth [42]. This connection was generalized by Frankl and Rödl [11, 35], who showed that the removal lemma for the complete k-uniform hypergraph with $k+1$ vertices implies Szemerédi's theorem for arithmetic progressions of length $k+1$. Moreover, Frankl and Rödl [11] verified such a removal lemma for $k = 3$ (see also [32] for the general removal lemma for 3-uniform hypergraphs) and Rödl and Skokan [40] for $k = 4$. The general result for k-uniform hypergraphs, based on generalizations of the regularity lemma and the local counting lemma for hypergraphs,

was obtained independently by Gowers [18] and by Nagle, Skokan, and authors [33, 39, 41]. Moreover Solymosi [44] and Tengan, Tokushige, and authors [38] showed that this result also implies multidimensional versions of Szemerédi's theorem first obtained by Furstenberg and Katznelson [13, 14].

Besides those extensions to hypergraphs, generalizations of Theorem 20 for graphs were proved by several authors. In particular, the regularity lemma of Alon, Fischer, Krivelevich, and M. Szegedy, Theorem 10, was introduced to prove the natural analog of the removal lemma for induced copies of F. In fact, the proof of this statement is already considerably more involved. Later, Alon and Shapira [4, 3] generalized those results by replacing the fixed graph F by a possibly infinite family of graphs $\mathcal{F}$. All those proofs relied on Theorem 10. The most general version, due to Alon and Shapira [3], states the following.

Theorem 21. *For every (possibly infinite) family of graphs $\mathcal{F}$ and every $\eta > 0$ there exist constants $c > 0$, $C > 0$, and n_0 such that the following holds. Suppose $G = (V, E)$ is a graph on $n \geq n_0$ vertices. If for every $\ell = 1, \ldots, C$ and every $F \in \mathcal{F}$ on ℓ vertices we have $N_F^*(G) < cn^\ell$, then there exists a graph $H = (V, E')$ on the same vertex set as G such that $|E \triangle E'| \leq \eta n^2$ and $N_F^*(H) = 0$ for every $F \in \mathcal{F}$.* ■

The proof of Theorem 21 is more involved and we will not present it here. The hypergraph extensions of Theorem 21 were obtained in [36].

Theorem 21 has interesting consequences in the area of *property testing*. Roughly speaking, it asserts that every graph G which is "far" (more than ηn^2 edges must be deleted or added) from some given hereditary property $\mathcal{A}$ (a property of graphs closed under isomorphism and vertex removal) must contain "many" ($cn^{|V(F)|}$) induced copies of some graph $F \notin \mathcal{A}$ of fixed size ($|V(F)| \leq C$). Consequently, a randomized algorithm can easily distinguish between graphs having $\mathcal{A}$ and those which are far from $\mathcal{A}$, provided $\mathcal{A}$ is decidable. One of the main questions in property testing, posed by Goldreich, Goldwasser, and Ron [16], concerns a natural characterization of properties allowing such a randomized algorithm. With respect to this question, the result of Alon and Shapira shows that all decidable, hereditary properties belong to that class (see [3] and [27] for more details).

An alternative proof of Theorem 21 was found by Lovász and B. Szegedy [28]. This new proof was based on the *limit approach* for sequences of dense graphs of those authors [29], which can be viewed as an infinitary iteration of Theorem 1. We will briefly explain this approach in the next section.

4. Graph Limits

In Section 2 we first introduced the (weak) regularity lemma of Frieze and Kannan and from an iterated version we deduced Szemerédi's regularity lemma and the (ε, r)-regularity lemma. Iterating Szemerédi's regularity lemma then resulted in the (strong) regularity lemma of Alon, Fischer, Krivelevich, and M. Szegedy, which was the key ingredient for the proof of Theorem 21.

It seems natural to further iterate any of those regularity lemmas. In fact, this was studied by Lovász and B. Szegedy [29]. Roughly speaking, those authors iterated the regularity lemma of Frieze and Kannan infinitely often. Below we will briefly outline some of their ideas. Note that due to the discussion above it does not matter which regularity lemma we iterate infinitely often, since we "pick up the other ones along the way".

Suppose $(G_i)_{i\in\mathbb{N}}$ is an infinite sequence of graphs with $\left|V(G_i)\right| \to \infty$ and $(\varepsilon_i)_{i\in\mathbb{N}}$ is a sequence of positive reals which tend to 0. Now we may apply Theorem 1 with ε_1 and $t_0 = 1$ to every sufficiently large graph G_i of the sequence. This way we obtain for every such graph G_i an ε_1-FK-partition $\mathcal{P}_{i,1}$ and a reduced graph $R_{i,1} = R_{G_i}(\mathcal{P}_{i,1})$. Note that all those partitions have at most $T_{\mathrm{FK}}(\varepsilon_1)$ parts. Hence, if we discretize the weights of the reduced graphs $R_{i,1}$ by quantities of up to ε_1, we note that there are only $\lceil 1/\varepsilon_1 \rceil^{\binom{T_{\mathrm{FK}}(\varepsilon_1)}{2}}$ different possible reduced graphs. Consequently, there exists a weighted graph R_1 with at most $T_{\mathrm{FK}}(\varepsilon_1)$ vertices such that $R_{i,1} = R_1$ for infinitely many choices $i \in \mathbb{N}$. In other words, there exists an infinite subsequence $\left(G_{i_j}\right)_{j\in\mathbb{N}}$ such that for every member there exists an ε_1-FK-partition, which yields R_1 as the reduced graph. We rename this sequence to $\left(G_i^1\right)_{i\in\mathbb{N}}$ and let $\left(\mathcal{P}_i^1\right)_{i\in\mathbb{N}}$ be the corresponding sequence of ε_1-FK-partitions.

We then repeat the above procedure with ε_2 for the infinite subsequence $\left(G_i^1\right)_{i\in\mathbb{N}}$, where the ε_2-FK-partitions should refine the ε_1-FK-partitions. This way we obtain a reduced graph R_2, an infinite subsequence $\left(G_i^2\right)_{i\in\mathbb{N}}$ of $\left(G_i^1\right)_{i\in\mathbb{N}}$, and a corresponding sequence of $\left(\mathcal{P}_i^2\right)_{i\in\mathbb{N}}$ of ε_2-FK-partitions. Repeating this step for every ε_j with $j \in \mathbb{N}$, we obtain a sequence of subsequences $\left(G_i^j\right)_{i\in\mathbb{N}}$ of graphs and a sequence of reduced graphs $(R_j)_{j\in\mathbb{N}}$. To avoid sequences of sequences of graphs we may pass to the diagonal sequence and let $(H_j)_{j\in\mathbb{N}} = \left(G_j^j\right)_{j\in\mathbb{N}}$ which is a subsequence of the original

sequence of graphs $(G_i)_{i\in\mathbb{N}}$. Moreover, for every $j \in \mathbb{N}$ and every $k = 1, \ldots, j$ let $\mathcal{Q}_j^k$ be the ε_k-FK-partition of H_j with reduced graph R_k.

Summarizing the above, we have argued that for every infinite sequence of graphs $(G_i)_{i\in\mathbb{N}}$ with $\left|V(G_i)\right| \to \infty$ and every sequence of positive reals $(\varepsilon_i)_{i\in\mathbb{N}}$ there exists a subsequence $(H_j)_{j\in\mathbb{N}}$ of $(G_i)_{i\in\mathbb{N}}$, and a sequence of reduced graphs $(R_j)_{j\in\mathbb{N}}$ such that

(*a*) for every $j \in \mathbb{N}$ and $k = 1, \ldots, j$ exist an ε_k-FK-partition $\mathcal{Q}_j^k$ of H_j with reduced graph R_k and

(*b*) for every $j \in \mathbb{N}$ and $k = 1, \ldots, j-1$ the partition $\mathcal{Q}_j^{k+1}$ refines $\mathcal{Q}_j^k$.

In some sense the graphs in the sequence $(H_j)_{j\in\mathbb{N}}$ become more and more similar, since they have almost identical FK-partitions for smaller and smaller ε. On the other hand, they may have very different sizes, which makes it hard to compare them directly. In order to circumvent that we may scale them all to the same size, by viewing them as functions on $[0,1]^2$. We will now make this more precise.

Let R_j be a reduced graph with t_j vertices. We split $[0,1]$ into t_j intervals $I_{j,1} \dot\cup \ldots \dot\cup I_{j,t_j} = [0,1]$ each of size $1/t_j$. We then define the symmetric, step-function $\hat{R}_j : [0,1]^2 \to [0,1]$ by setting

$$\hat{R}_j(x,y) = \begin{cases} w_{R_j}(k,\ell), & \text{if } (x,y) \text{ belongs to the interior of } I_{j,k} \times I_{j,\ell}, \\ 0, & \text{otherwise.} \end{cases}$$

Recall that those reduced graphs R_j came from refining partitions. It will be important to impose the same structure for the implicit partitions $I_{j,1} \dot\cup \ldots \dot\cup I_{j,t_j} = [0,1]$ for $j \in \mathbb{N}$.

In particular, we recall that every vertex of R_j represents a vertex class of $\mathcal{Q}_{j'}^j$, for every $j' \geq j$ and those vertex classes were split into t_{j+1}/t_j vertex classes in $\mathcal{Q}_{j'}^{j+1}$ for every $j' > j$. In other words, we may view every vertex of R_j of being split into t_{j+1}/t_j vertices of R_{j+1}. Therefore, we require that for every j the partition $I_{j+1,1} \dot\cup \ldots \dot\cup I_{j+1,t_{j+1}}$ refines $I_{j,1} \dot\cup \ldots \dot\cup I_{j,t_j}$ in the same way, i.e., the t_{j+1}/t_j intervals $I_{j+1,x}$ representing vertices of R_{j+1} which came from a given vertex y of R_j form a partition of $I_{j,y}$.

This way we view the sequence of reduced graphs $(R_j)_{j\in\mathbb{N}}$ as a sequence of symmetric step-functions from $[0,1]^2 \to [0,1]$. Similarly, we may view the sequence of graphs $(H_j)_{j\in\mathbb{N}}$ as a sequence of symmetric step-functions

from $[0,1]^2 \to \{0,1\}$. Here for a graph H_j on n_j vertices we split $[0,1]$ into n_j intervals $J_{j,1} \dot\cup \ldots \dot\cup J_{j,n_j} = [0,1]$ (identified by the vertices of H_j) and we set

$$\hat{H}_j(x,y) = \begin{cases} 1, & \text{if } (x,y) \text{ belongs to the interior of } J_{j,u} \times J_{j,v} \\ & \text{and } \{u,v\} \in E(H_j), \\ 0, & \text{otherwise.} \end{cases}$$

Here we require that the labeling of the vertices of H_j is "consistent", i.e., if u is a vertex contained in the k-th vertex class of the fixed ε_j-FK-partition of H_j, then we impose that $J_{j,u} \subseteq I_{j,k}$.

After this embedding we can rewrite the property that R_j is the reduced graph of an ε_j-FK-partition of H_j, by

$$\sup_{U \subseteq [0,1]} \left| \int_{U \times U} \hat{H}_j(x,y) - \hat{R}_j(x,y)\, \mathrm{d}x\, \mathrm{d}y \right| \leq \hat{\varepsilon}_j, \tag{21}$$

for some $\hat{\varepsilon}_j$ which tends to 0 as ε_j tends to 0, where (here and below) the supremum is taken over all (Lebesgue) measurable subset $U \subseteq [0,1]$. (Note that $\hat{H}_j$ and $\hat{R}_j$ are piecewise linear and, hence, (Lebesgue) measurable on $[0,1]^2$.) Moreover, we can rephrase the global counting lemma, Theorem 16: Let F be a graph with $V(F) = [\ell]$ and let j be sufficiently large (so that ε_j is sufficiently small). Then

$$\frac{N_F(H_j)}{n_j^\ell} = \int_{(x_1,\ldots,x_\ell) \in [0,1]^\ell} \prod_{\{p,q\} \in E(F)} \hat{R}_j(x_p, x_q)\, \mathrm{d}x_1 \ldots \mathrm{d}x_\ell \pm \hat{\gamma}_j, \tag{22}$$

where for fixed F we have $\hat{\gamma}_j \to 0$ as $\varepsilon_j \to 0$.

It was proved by Lovász and B. Szegedy in [29] that, due to property (b) above, the sequence $(\hat{R}_j)_{j \in \mathbb{N}}$ converges almost everywhere to a measurable, symmetric function $\hat{R} : [0,1]^2 \to [0,1]$ and that (21) and (22) stay valid in the limit. The function $\hat{R}$ is called the limit of the sequence $(H_j)_{j \in \mathbb{N}}$.

Theorem 22. *For every sequence of graphs $(G_i)_{i \in \mathbb{N}}$ with $|V(G_i)| \to \infty$ there exists a subsequence $(H_j)_{j \in \mathbb{N}}$ and a sequence of reduced graphs $(R_j)_{j \in \mathbb{N}}$, and a measurable, symmetric function $\hat{R} : [0,1]^2 \to [0,1]$ such that*

(i) $\hat{R}_j$ converges pointwise almost everywhere to $\hat{R}$,

(*ii*)

$$\lim_{j\to\infty} \sup_{U\subseteq[0,1]} \left| \int_{U\times U} \hat{H}_j(x,y) - \hat{R}(x,y)\,\mathrm{d}x\,\mathrm{d}y \right| = 0,$$

and

(*iii*) *for every* $\ell \in \mathbb{N}$ *and every graph* F *with* $V(F) = [\ell]$

$$\lim_{j\to\infty} \frac{N_F(H_j)}{n_j^\ell} = \int_{(x_1,\ldots,x_\ell)\in[0,1]^\ell} \prod_{\{p,q\}\in E(F)} \hat{R}(x_p,x_q)\,\mathrm{d}x_1\ldots\mathrm{d}x_\ell. \qquad \blacksquare$$

The proof of Theorem 22 indicated above, essentially follows the lines of the proof of the implication $(a) \Rightarrow (b)$ of Theorem 2.2 in [29] (see Lemma 5.1 and 5.2 in [29]). Based on Theorem 22 Lovász and B. Szegedy [28] gave a different and conceptually simpler proof of Theorem 21. The proof of the generalization of Theorem 21 to k-uniform hypergraphs in [36] followed similar ideas (see also [5]).

References

[1] N. Alon, A. Coja-Oghlan, H. Hàn, M. Kang, V. Rödl and M. Schacht, Quasi-randomness and algorithmic regularity for graphs with general degree distributions, *SIAM J. Comput.*, to appear.

[2] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, *Combinatorica,* **20** (2000), no. 4, 451–476.

[3] N. Alon and A. Shapira, A characterization of the (natural) graph properties testable with one-sided error, *SIAM J. Comput.,* **37** (2008), no. 6, 1703–1727.

[4] N. Alon and A. Shapira, Every monotone graph property is testable, *SIAM J. Comput.,* **38** (2008), no. 2, 505–522.

[5] T. Austin and T. Tao, *On the testability and repair of hereditary hypergraph properties,* submitted.

[6] W. G. Brown, P. Erdős and V. T. Sós, *Some extremal problems on r-graphs,* New directions in the theory of graphs (Proc. Third Ann Arbor Conf., Univ. Michigan, Ann Arbor, Mich, 1971), Academic Press, New York, 1973, pp. 53–63.

[7] V. Chvátal, V. Rödl, E. Szemerédi and W. T. Trotter, Jr., The Ramsey number of a graph with bounded maximum degree, *J. Combin. Theory Ser. B,* **34** (1983), no. 3, 239–243.

[8] R. A. Duke, H. Lefmann and V. Rödl, A fast approximation algorithm for computing the frequencies of subgraphs in a given graph, *SIAM J. Comput.,* **24** (1995), no. 3, 598–620.

[9] P. Erdős and P. Turán, On some sequences of integers, *J. London. Math. Soc.,* **11** (1936), 261–264.

[10] P. Erdős, P. Frankl and V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs Combin.,* **2** (1986), no. 2, 113–121.

[11] P. Frankl and V. Rödl, Extremal problems on set systems, *Random Structures Algorithms,* **20** (2002), no. 2, 131–164.

[12] A. Frieze and R. Kannan, Quick approximation to matrices and applications, *Combinatorica,* **19** (1999), no. 2, 175–220.

[13] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations, *J. Analyse Math.,* **34** (1978), 275–291 (1979).

[14] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for IP-systems and combinatorial theory, *J. Analyse Math.,* **45** (1985), 117–168.

[15] S. Gerke and A. Steger, *The sparse regularity lemma and its applications,* Surveys in combinatorics 2005, London Math. Soc. Lecture Note Ser., vol. 327, Cambridge Univ. Press, Cambridge, 2005, pp. 227–258.

[16] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, *J. ACM,* **45** (1998), no. 4, 653–750.

[17] W. T. Gowers, Lower bounds of tower type for Szemerédi's uniformity lemma, *Geom. Funct. Anal.,* **7** (1997), no. 2, 322–337.

[18] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, *Ann. of Math.* (2), **166** (2007), no. 3, 897–946.

[19] P. E. Haxell, Y. Kohayakawa and T. Łuczak, The induced size-Ramsey number of cycles, *Combin. Probab. Comput.,* **4** (1995), no. 3, 217–239.

[20] P. E. Haxell, Y. Kohayakawa and T. Łuczak, Turán's extremal problem in random graphs: forbidding even cycles, *J. Combin. Theory Ser. B,* **64** (1995), no. 2, 273–287.

[21] P. E. Haxell, Y. Kohayakawa and T. Łuczak, Turán's extremal problem in random graphs: forbidding odd cycles, *Combinatorica,* **16** (1996), no. 1, 107–122.

[22] Y. Kohayakawa, *Szemerédi's regularity lemma for sparse graphs,* Foundations of computational mathematics (Rio de Janeiro, 1997), pp. 216–230.

[23] Y. Kohayakawa and V. Rödl, *Szemerédi's regularity lemma and quasi-randomness,* Recent advances in algorithms and combinatorics, CMS Books Math./Ouvrages Math. SMC, vol. 11, Springer, New York, 2003, pp. 289–351.

[24] J. Komlós, G. N. Sárközy and E. Szemerédi, Blow-up lemma, *Combinatorica,* **17** (1997), no. 1, 109–123.

[25] J. Komlós, A. Shokoufandeh, M. Simonovits and E. Szemerédi, *The regularity lemma and its applications in graph theory,* Theoretical aspects of computer science (Tehran, 2000), Lecture Notes in Comput. Sci., vol. 2292, Springer, Berlin, 2002, pp. 84–112.

[26] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its applications in graph theory,* Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud., vol. 2, János Bolyai Math. Soc., Budapest, 1996, pp. 295–352.

[27] L. Lovász and B. Szegedy, Testing properties of graphs and functions, *Israel J. Math.,* to appear.

[28] L. Lovász and B. Szegedy, *Graph limits and testing hereditary graph properties,* Tech. Report MSR-TR-2005-110, Microsoft Research, 2005.

[29] L. Lovász and B. Szegedy, Limits of dense graph sequences, *J. Combin. Theory Ser. B,* **96** (2006), no. 6, 933–957.

[30] L. Lovász and B. Szegedy, Szemerédi's lemma for the analyst, *Geom. Funct. Anal.,* **17** (2007), no. 1, 252–270.

[31] T. Łuczak, *Randomness and regularity,* International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 899–909.

[32] B. Nagle and V. Rödl, Regularity properties for triple systems, *Random Structures Algorithms,* **23** (2003), no. 3, 264–332.

[33] B. Nagle, V. Rödl and M. Schacht, The counting lemma for regular k-uniform hypergraphs, *Random Structures Algorithms,* **28** (2006), no. 2, 113–179.

[34] V. Rödl, unpublished.

[35] V. Rödl, *Some developments in Ramsey theory,* Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990) (Tokyo), Math. Soc. Japan, 1991, pp. 1455–1466.

[36] V. Rödl and M. Schacht, Generalizations of the removal lemma, *Combinatorica,* **29** (2009), no. 4, 467–501.

[37] V. Rödl and M. Schacht, Regular partitions of hypergraphs: regularity lemmas, *Combin. Probab. Comput.,* **16** (2007), no. 6, 833–885.

[38] V. Rödl, M. Schacht, E. Tengan and N. Tokushige, Density theorems and extremal hypergraph problems, *Israel J. Math.,* **152** (2006), 371–380.

[39] V. Rödl and J. Skokan, Regularity lemma for k-uniform hypergraphs, *Random Structures Algorithms,* **25** (2004), no. 1, 1–42.

[40] V. Rödl and J. Skokan, Counting subgraphs in quasi-random 4-uniform hypergraphs, *Random Structures Algorithms,* **26** (2005), no. 1–2, 160–203.

[41] V. Rödl and J. Skokan, Applications of the regularity lemma for uniform hypergraphs, *Random Structures Algorithms,* **28** (2006), no. 2, 180–194.

[42] K. F. Roth, On certain sets of integers, *J. London Math. Soc.,* **28** (1953), 104–109.

[43] I. Z. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles,* Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II, Colloq. Math. Soc. János Bolyai, vol. 18, North-Holland, Amsterdam, 1978, pp. 939–945.

[44] J. Solymosi, A note on a question of Erdős and Graham, *Combin. Probab. Comput.,* **13** (2004), no. 2, 263–267.

[45] V. T. Sós, P. Erdős and W. G. Brown, On the existence of triangulated spheres in 3-graphs, and related problems, *Period. Math. Hungar.*, **3** (1973), no. 3–4, 221–228.

[46] E. Szemerédi, On graphs containing no complete subgraph with 4 vertices, *Mat. Lapok*, **23** (1972), 113–116 (1973).

[47] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.*, **27** (1975), 199–245, Collection of articles in memory of Jurii Vladimirovič Linnik.

[48] E. Szemerédi, *Regular partitions of graphs,* Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), Colloq. Internat. CNRS, vol. 260, CNRS, Paris, 1978, pp. 399–401.

[49] T. Tao, Szemerédi's regularity lemma revisited, *Contrib. Discrete Math.*, **1** (2006), no. 1, 8–28 (electronic).

Vojtěch Rödl
Department of Mathematics and Computer Science,
Emory University,
Atlanta, GA 30322,
U.S.A.
e-mail: rodl@mathcs.emory.edu

Mathias Schacht
Institut für Informatik,
Humboldt-Universität zu Berlin,
Unter den Linden 6,
D-10099 Berlin,
Germany
e-mail:
schacht@informatik.hu-berlin.de

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 327–336.

Edge Coloring Models as Singular Vertex Coloring Models

BALÁZS SZEGEDY

Dedicated to the 60th birthday of László Lovász

In this paper we study an interesting class of graph parameters. These parameters arise both from edge coloring models and from limits of vertex coloring models with a fixed number of vertices. Their vertex connection matrices have exponentially bounded rank growth.

1. Introduction

Vertex coloring models (resp. edge coloring models) were introduced by Freedman, Lovász and Schrijver. The main idea is that any graph G can be imagined as a particle system where the particles are represented by the vertices (resp. edges). The particles can have a finite number of states (called colors) and the states interact with each other along the edges (resp. vertices).

In a vertex coloring model every state and interaction between two states has a weight. This data can be best represented as a weighted graph H on the color set. The partition function of the vertex coloring model is defined as the homomorphism number of G into this weighted graph H (see [1]).

Freedman, Lovász and Schrijver in [1] give a characterization of the partition functions of vertex coloring models (with real weights) in terms of the so called connection matrices $M(p,k)$ that can be defined for an arbitrary graph parameter p and number k. They show that if $\mathrm{rk}\left(M(p,k)\right)$ grows at most exponentially with k and $M(p,k)$ is positive semidefinite for

every k then there is a weighted graph H (with positive vertex weights and real edge weights) such that p is the homomorphism function into H. A similar characterization of the partition functions of edge coloring models was obtained by the author [2].

The main motivation of this paper is an interesting observation of L. Lovász. Let $p(G)$ denote the number of perfect matchings in G. Lovász noticed that the parameter p is similar to a homomorphism number in the sense that $\mathrm{rk}\left(M(p,k)\right)$ has exponential rank growth 2^k but $M(p,k)$ is not positive semidefinite. However it is not the homomorphism number into any weighted graph. Note that p is the partition function of an edge coloring model.

Motivated by this observation the author found the following curious construction. Let H_t be the weighted graph on two vertices with edge weights

$$\begin{pmatrix} 1+t^2 & 1 \\ 1 & 1 \end{pmatrix}$$

and vertex weights

$$\left(\frac{1}{t}, \frac{-1}{t}\right).$$

Then

$$p(G) = \lim_{t \to 0} \hom(G, H_t)$$

for every G. In particular this formula explains the exponential rank growth because $M(p,k)$ is the limit of matrices with rank 2^k. The conclusion of the present paper is that a wide class of graph parameters that are coming from edge coloring models can be represented in a similar way. The generalized vertex coloring models will be called singular vertex coloring models. The main theorem says that if an edge coloring model satisfies a certain finiteness condition (which seems to be satisfied by most reasonable parameters) then its partition function can be represented by a singular vertex coloring model. Note that this result is complemented by a result in [2] saying that partition functions of vertex coloring models can be represented by partition functions of edge coloring models with complex values.

Acknowledgement. I am very grateful to László Lovász for mentoring me throughout my post-doc years (2003–2005) when these results were obtained.

2. Finite Rank Functions on Polynomial Rings

We will need some basic facts from commutative algebra.

Let $R = \mathbb{C}[x_1, x_2, \ldots, x_n]$ be the polynomial ring in n variables and let $\mathcal{F} : R \to \mathbb{C}$ be a $\mathbb{C}$-linear function. We define the symmetric bilinear form $b : R \times R \to \mathbb{C}$ by $b(p,q) = \mathcal{F}(pq)$. Let I_b denote the set of polynomials p such that $b(p,q) = 0$ for every $q \in R$. It is clear that I_b is an ideal in R. We say that $\mathcal{F}$ has **finite rank** if $\dim(R/I_b) < \infty$. In general $\dim(R/I_b)$ will be called the rank of $\mathcal{F}$. We say that $\mathcal{F}$ is **semisimple** if I_b is a finite co-dimensional radical ideal in R. A function $\mathcal{F}$ is semisimple if and only if $R/I_b \equiv \mathbb{C}^k$ for some natural number k. Finite rank functions are forming a subspace of the dual space of R and semisimple functions are forming a subspace of all finite rank functions. The next classical lemma ([3], [4]) provides generating sets for these spaces.

Lemma 2.1 (Generating finite rank functions). *Let m be a multiset of the variables $x_1, x_2, \ldots, x_n$ and $v \in \mathbb{C}^n$ be a vector. Then the functions $\mathcal{F}_{m,v}$ defined by*

$$\mathcal{F}_{m,v}(p) = \partial_m p(v)$$

are all finite rank functions and they linearly span the space of finite rank functions. Furthermore the space of semisimple functions is generated by the functions of the form $\mathcal{F}_v(p) = p(v)$.

The following lemma is a corollary of the previous one.

Lemma 2.2 (Approximation of finite rank functions). *Let $\mathcal{F}$ be an arbitrary finite rank function on R. Then there are two lists of vectors $\{v_i\}_{i=1}^d$ and $\{w_i\}_{i=1}^d$ with $v_i, w_i \in \mathbb{C}^n$ and coefficients $\{c_i/t^{r_i}\}_{i=1}^d$ with $c_i \in \mathbb{C}$ and $r_i \in \mathbb{N}$ such that*

$$\mathcal{F}(p) = \lim_{t \to 0} \sum_{i=1}^{d} \frac{c_i}{t^{r_i}} p(v_i + t w_i)$$

for every polynomial R.

Proof. By Lemma 2.1 it is enough to prove the lemma for functions of the for $\mathcal{F}_{m,v}$. Let e_i denote the vector in $\mathbb{C}^n$ whose i-th coordinate is 1 and all other coordinates are 0. Let furthermore m_i denote the multiplicity of x_i in m for $1 \leq i \leq n$. Then it is classical that

$$(1) \quad \mathcal{F}_{m,v}(p) = \lim_{t \to 0} \frac{1}{t^b} \sum_{a_1=0}^{m_1} \sum_{a_2=0}^{m_2} \cdots \sum_{a_n=0}^{m_n} p\left(v + t \sum_{k=1}^{n} a_k e_k\right) \prod_{j=1}^{n} (-1)^{a_j} \binom{m_j}{a_j}$$

where $b = \sum_{k=1}^{n} m_k$. ■

3. Edge Coloring Models of Finite Rank

The so-called edge coloring models were introduced by Freedman, Lovász and Schrijver. In this chapter we introduce a special class of edge coloring models with interesting properties.

An edge coloring model is represented by a function $f : \mathbb{N}^n \to \mathbb{C}$. Let $\mathcal{C} = \{c_1, c_2, \ldots, c_n\}$ be a set whose elements are called **colors.** By abusing the notation, for a multiset S of colors we define $f(S)$ as the evaluation of f on the vector $(a_1, a_2, \ldots, a_n)$ where a_i denotes the multiplicity of c_i in S. The partition function p_f of the edge coloring model represented by f is a graph parameter defined by the following formula.

$$p_f(G) = \sum_{\psi : E(G) \to \mathcal{C}} \prod_{v \in V(G)} f\big(S(v, \psi)\big)$$

where $S(v, \psi)$ denotes the multiset of values of ψ on the edges adjacent to v.

Let us introduce an infinite matrix M_f whose rows and columns are indexed by the elements of $\mathbb{N}^n$ and $M_f(v, w) = f(v+w)$ for every $v, w \in \mathbb{N}^n$. The edge coloring model corresponding to f is called **finite rank edge coloring model** if the rank of M_f is finite.

The main method in this paper relies on the fact that an edge coloring model $f : \mathbb{N}^n \to \mathbb{C}$ can be turned into a linear function $\mathcal{F} : \mathbb{C}[x_1, x_2, \ldots, x_n] \to \mathbb{C}$ in the following way. We define $\mathcal{F}(x_1^{a_1} x_2^{a_2} \ldots x_n^{a_n})$ as $f(a_1, a_2, \ldots, a_n)$ and then we extend it linearly to all polynomials. We call $\mathcal{F}$ the **polynomial extension** of f. The main observation is that f is of finite rank if and only if $\mathcal{F}$ is of finite rank. This can be seen from the fact that the semigroup $\mathbb{N}^n$ is naturally represented as the multiplicative semigroup of monomials in $\mathbb{C}[x_1, x_2, \ldots, x_n]$. We will say that f is a **semisimple edge coloring model** if the corresponding function $\mathcal{F}$ is semisimple.

4. Singular Vertex Coloring Models

Similarly to edge coloring models, the so called vertex coloring models were first introduced by Freedman, Lovász and Schrijver in the paper [1]. We start with the definition of vertex coloring models and connection matrices and then we introduce a generalization of vertex coloring models.

Let H be a complete graph on the vertex set $[n] = \{1, 2, \ldots, n\}$ with weights α_i on the vertices and weights $\beta_{i,j}$ on the edges. The homomorphism number $\hom(G, H)$ for a simple graph G is defined as follows.

$$\hom(G,H) = \sum_{\phi: V(G)\to[n]} \prod_{v\in V(G)} \alpha_{\phi(v)} \prod_{(v,w)\in E(G)} \beta_{\phi(v),\phi(w)}.$$

If H is fixed then the graph parameter $t_H(G) = \hom(G, H)$ is called the partition function of the vertex coloring model corresponding to H.

Let $\mathcal{G}_k$ denote the set of simple graphs with k vertices labeled by the set $[k]$. We denote by $\mathcal{G} = \mathcal{G}_0$ the set of graphs in which no vertices are labeled. The set $\mathcal{G}_k$ has a commutative semigroup structure with respect to the natural gluing operation along the labeled edges. For two elements $a, b \in \mathcal{G}_k$ their on product in this semigroup is denoted by ab. Let $p : \mathcal{G} \to \mathbb{C}$ be an arbitrary graph parameter. The k-th **connection matrix** of p is an infinite matrix $M(p, k)$ of the form $\mathbb{C}^{\mathcal{G}_k \times \mathcal{G}_k}$ such that $M(p, k)(a, b) = p(ab)$. Freedman, Lovász and Schrijver observed that if H is a fixed weighted graph on n vertices then

$$\operatorname{rk}\left(M(t_H, k)\right) \leq n^k.$$

In other words, the rank growth of the connection matrices is at most exponential. Note that the above formula for $k = 0$ implies that t_H is multiplicative when taking disjoint unions of graphs.

Now we are ready to define the generalization of vertex models. Let H be as above with the modification that the edge weights are in $\mathbb{C}[t]$ and the vertex weights are of the form c/t^s where $c \in \mathbb{C}$ and $s \in \mathbb{N}$. We say that H gives rise to a **singular vertex coloring model** if

$$p(G) = \lim_{t\to 0} \hom(G, H) \in \mathbb{C}$$

for every $G \in \mathcal{G}$. The graph parameter p is called the partition function of the singular vertex model. Every vertex coloring model is a singular vertex coloring model in a way that the edge and vertex weights don't depend on t. The following lemma follows trivially from the fact that $M(p, k) = \lim_{t\to 0} M(t_H, k)$ and that $\operatorname{rk}\left(M(t_H, k)\right) \leq n^k$.

Lemma 4.1 (Exponential rank growth). *Let H be a singular vertex coloring model. Then its partition function p satisfies* $\mathrm{rk}\big(M(p,k)\big) \leq n^k$ *for every $k \in \mathbb{N}$ where n is the number of vertices in H.*

The main theorem in this paper is the following.

Theorem 1 (Finite rank edge models as singular vertex models). *The partition function of any finite rank edge coloring model is the same as the partition function of some singular vertex coloring model.*

The next curious corollary says that the vertex connection matrices of partition functions of finite rank edge coloring models have exponential rank grows. Note that the edge connection matrices of edge coloring models have exponential rank grow.

Corollary 4.1 (Exponential rank growth II). *The partition function p_f of any finite rank edge coloring model satisfies* $\mathrm{rk}\big(M(t_f,k)\big) \leq c^k$ *for every $k \in \mathbb{N}$ and constant c depending on f.*

For two vectors $v = (a_1, a_2, \ldots, a_n)$ and $w = (b_1, b_2, \ldots, b_n)$ in $\mathbb{C}^n$ let $\langle v, w\rangle$ denote the sum $a_1b_1 + a_2b_2 + \cdots + a_nb_n$. For proving theorem 1 we will make use of the next lemma.

Lemma 4.2 (Semisimple edge models as vertex models). *Let $f : \mathbb{N}^n \to \mathbb{C}$ be a semisimple edge coloring model of the form*

$$\mathcal{F}(p) = \sum_{i=1}^{d} \lambda_i p(v_i)$$

where $\lambda_i \in \mathbb{C}$ and $v_i \in \mathbb{C}^n$. Then the partition function of f is the same as the homomorphism number into the weighted graph H on the vertex set $[d]$ with weights

$$\alpha_i = \lambda_i,\ \beta_{i,j} = \langle v_i, v_j\rangle \quad 1 \leq i,j \leq d.$$

Proof. Let $f_i : \mathbb{N}^n \to \mathbb{C}$ denote the function whose value on $(a_1, a_2, \ldots, a_n)$ is the substitution of v_i into the monomial $x_1^{a_1}x_2^{a_2}\ldots, x_n^{a_n}$. We have that $f = \sum_{i=1}^{d} \lambda_i f_i$. Let $\hat{f}_i : \mathcal{C} \to \mathbb{C}$ denote the function whose value on c_j is the $j-th$ component of v_i. Then

$$p_f(G) = \sum_{\psi : E(G)\to\mathcal{C}} \prod_{v\in V(G)} \sum_{i=1}^{d} \lambda_i f_i\big(S(v,\psi)\big)$$

$$= \sum_{\psi : E(G)\to\mathcal{C}} \sum_{\phi : V(G)\to[d]} \prod_{v\in V(G)} \lambda_{\phi(v)} f_{\phi(v)}\big(S(v,\psi)\big)$$

$$= \sum_{\phi: V(G)\to[d]} \sum_{\psi: E(G)\to\mathcal{C}} \left(\prod_{v\in V(G)} \lambda_{\phi(v)} \right)$$

$$\times \prod_{e=(u,w)\in E(G)} \hat{f}_{\phi(u)}(\psi(e))\, \hat{f}_{\phi(w)}(\psi(e))$$

$$= \sum_{\phi: V(G)\to[d]} \left(\prod_{v\in V(G)} \lambda_{\phi(v)} \right) \prod_{(u,w)\in E(G)} \sum_{k=1}^{n} \hat{f}_{\phi(u)}(c_k) \hat{f}_{\phi(w)}(c_k)$$

$$= \sum_{\phi: V(G)\to[d]} \left(\prod_{v\in V(G)} \lambda_{\phi(v)} \right) \prod_{(u,w)\in E(G)} \langle v_{\phi(u)}, u_{\phi(w)} \rangle . \qquad \blacksquare$$

Now we are ready to prove the main theorem.

Proof of Theorem 1: Let f be a finite rank edge coloring model and $\mathcal{F}$ be its polynomial extension. Lemma 2.2 yields a one parameter family of semisimple functions $\mathcal{F}_t$ such that $\lim_{t\to 0} \mathcal{F}_t(q) = \mathcal{F}(q)$ for every n variable polynomial q. Let $f(t)$ be the edge coloring model obtained form $\mathcal{F}_t$ by restricting it to monomials. We have that $\lim_{t\to 0} f(t)(a_1, a_2, \ldots, a_n) = f(a_1, a_2, \ldots, a_n)$ for every $(a_1, a_2, \ldots, a_n) \in \mathbb{C}^n$ and so $\lim_{t\to 0} p_{f(t)}(G) = p_f(g)$ for every simple graph G. Now using Lemma 4.2 and formula (1) the proof is complete.

5. Hybrid Models and Directed Edge Models

In this chapter we discuss a common generalization of vertex and edge coloring models that were first introduced by M. Freedman. We call them **hybrid models.** A **half edge** in a graph G is an incident pair (v, e) of a vertex v and edge e. We denote by $E^*(G)$ the set of half edges in G. Roughly speaking, in a hybrid model a graph is imagined as a particle system in which the half edges can have different states (or colors) that interact with each other both at the vertices and along the edges. The precise definition is the following.

Assume that beside a function $f : \mathbb{N}^n \to \mathbb{C}$ a symmetric matrix $A \in \mathbb{C}^{n\times n}$ is given whose rows and columns are indexed by the elements of the color set $\mathcal{C}$. The pair (f, A) is called a hybrid model. Similarly to edge coloring models, we say that (f, A) is a **finite rank hybrid model** if the

polynomial extension $\mathcal{F}$ of f has finite rank. The partition function p of the hybrid model (f, A) is given by

$$p(G) = \sum_{\psi : E^*(G) \to \mathcal{C}} \left(\prod_{v \in V(G)} f\big(S(v,\psi)\big) \right) \prod_{e=(v,w) \in E(G)} A\big(\psi(v,e), \psi(w,e)\big).$$

where $S(v,\psi)$ denotes the multiset of the values of ψ on the half edges incident to v.

It can be derived from the results in [2] that the partition function of any hybrid model is always equal to the partition function of some edge coloring model. So in the sense of partition functions they are not more general. However some combinarial parameters arise more naturally from hybrid models than from pure edge coloring models. For example a variant of the edge coloring model in which edges have colors and directions can be represented as hybrid models.

An important feature of hybrid models is that they have finite rank if and only if the representing edge coloring model has finite rank. This allows one to use the machinery developed in this paper for hybrid models.

6. Examples and Remarks

Finite rank edge coloring models are very common in the sense that edge coloring models with "nice" combinatorial meaning are typically of this type. We list a few conditions that give a lot of finite rank models. Let $f, g : \mathbb{N}^n \to \mathbb{C}$ be edge coloring models.

1.) If f takes only finitely many non 0 values then it has finite rank.

2.) If f and g have finite rank then cf, $f+g$ and fg have finite rank where $c \in \mathbb{C}$ and the operations are performed pointwise on $\mathbb{N}^n$. Together with the fact that constant functions have finite rank, it implies that any polynomial expression of finite rank functions is again of finite rank.

3.) If f has finite rank and $h : \mathbb{N}^{n+m} \to \mathbb{C}$ is obtained from f in the way that $h(a_1, a_2, \ldots, a_{n+m}) = f(a_1, a_2, \ldots, a_n)$ then h has finite rank.

4.) If g is obtained form f by permuting the coordinates in $\mathbb{N}^n$ and f has finite rank then g has finite rank.

5.) The coordinate functions $c_i(a_1, a_2, \ldots, a_n) = a_i$ have all finite rank.

6.) Exponentials of the coordinate functions l^{c_i} with a fixed $l \in \mathbb{C}$ have all finite rank.

Now here are some examples:

Number of perfect matchings. Let f be the two variable edge coloring model with $f(a,b) = 1$ if $a = 1$ and 0 if $a \neq 1$. Conditions 1. and 3. show that f has finite rank. The corresponding partition function is the number of perfect matchings. The polynomial extension $\mathcal{F}$ of f is equal to $\mathcal{F}_{x_1,(0,1)}$. Thus the semisimple approximation of $\mathcal{F}$ given by Lemma 2.2 is $\mathcal{F}_t(q) = \frac{1}{t}\big(q(t,1) - q(0,1)\big)$. This implies by Lemma 4.2 that the singular vertex coloring model has two vertices and furthermore $\alpha_1 = 1/t$, $\alpha_2 = -1/t$, $\beta_{1,1} = t^2 + 1$, $\beta_{1,2} = \beta_{2,1} = \beta_{2,2} = 1$.

Being three regular. Let f be the one variable edge coloring model with $f(a) = 1$ of $a = 3$ and $f(a) = 0$ if $a \neq 3$. This is of finite rank by condition 1. The partition function of f gives 1 if and only if G is 3-regular. In this case $\mathcal{F}_t(q) = \frac{1}{6t^3}\big(q(3t) - 3q(2t) + 3q(t) - q(0)\big)$ and so the singular vertex coloring model has four points. The vertex weights are $-1/6t^3$, $3/6t^3$, $-3/6t^3$, $1/6t^3$ and the edge weights are $\beta_{i,j} = (i-1)(j-1)t^2$.

Number of Eulerian subgraphs. Let f be the two variable edge coloring model with $f(a,b) = 1$ if a is even and 0 if a is odd. Conditions 2., 3. and 6. together show that f is of finite rank. An interesting fact is that f is semisimple so it can be represented by a normal vertex coloring model (this was first observed by L. Lovász). Further details are left to the reader.

Other examples. Many examples for edge coloring models listed in [2] are of finite rank. This includes the number of fully packed loop configurations, the number of matchings, the number of d-regular subgraphs, the number of proper edge colorings with d-colors and the permanent of the adjacency matrix.

We finish the paper with an open question.

Question. *Is there any graph parameter p with exponentially bounded rank growth $\big(\mathrm{rk}\big(M(p,k)\big) \leq c^k\big)$ which is not the partition function of a finite rank edge coloring model?*

References

[1] M. Freedman, L. Lovász and A. Schrijver, Reflection positivity, rank connectivity, and homomorphism of graphs, *J. Amer. Math. Soc.*, **20** (2007), no. 1, 37–51.

[2] B. Szegedy, Edge coloring models and reflection positivity, *J. Amer. Math. Soc.*, **20** (2007), no. 4, 969–988.

[3] W. Gröbner, *Algebraische Geometrie II,* Bibliographisches Institut, Mannheim, 1970.

[4] I. Kaplansky, *Commutative Rings,* The University of Chicago Press, Chicago, 1974.

Balázs Szegedy
University of Toronto
Dept. of Mathematics
40 St George St.
Toronto, ON
M5S 2E4, Canada

e-mail: szegedyb@gmail.com

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 337–353.

List Total Weighting of Graphs

TSAI-LIEN WONG*, DAQING YANG† and XUDING ZHU‡

A graph $G = (V, E)$ is (k, k')-total weight choosable if the following is true: For any (k, k')-total list assignment L that assigns to each vertex v a set $L(v)$ of k real numbers as permissible weights, and assigns to each edge e a set $L(e)$ of k' real numbers as permissible weights, there is a proper L-total weighting, i.e., a mapping $f : V \cup E \to \mathbb{R}$ such that $f(y) \in L(y)$ for each $y \in V \cup E$, and for any two adjacent vertices u and v, $\sum_{e \in E(u)} f(e) + f(u) \neq \sum_{e \in E(v)} f(e) + f(v)$. This paper introduces a method, the max-min weighting method, for finding proper L-total weightings of graphs. Using this method, we prove that complete multipartite graphs of the form $K_{n,m,1,1,\ldots,1}$ are $(2, 2)$-total weight choosable and complete bipartite graphs other than K_2 are $(1, 2)$-total weight choosable.

1. Introduction

Suppose $G = (V, E)$ is a graph. For a vertex v of G, let $E(v)$ be the set of edges of G incident to v, and let $E^*(v) = E(v) \cup \{v\}$. An *edge weighting* of G is a mapping that assigns to each edge e of G a real number $f(e)$. An edge weighting f induces a vertex colouring $\varphi_f : V \to \mathbb{R}$ of G, defined as $\varphi_f(v) = \sum_{e \in E(v)} f(e)$. We say f is a *proper edge weighting* if the induced vertex colouring φ_f is proper, i.e., for any edge uv of G, $\varphi_f(u) \neq \varphi_f(v)$. The study of edge weighting was initiated by Karoński, Łuczak and Thomason [9]. They made an interesting conjecture: Every connected graphs $G \neq K_2$ has a proper edge weighting f such that $f(e) \in \{1, 2, 3\}$ for every edge e. This conjecture is still open. It was shown in [3] that the conjecture would

*Grant number: NSC97-2115-M-110-004-MY2.

†Supported in part by NSFC under grant 10771035 and 10931003 grant SX2006-42 of colleges of Fujian.

‡Grant number: NSC97-2115-M-110-008-MY3.

be true if the set $\{1,2,3\}$ is replaced by $\{1,2,\ldots,30\}$. The result was improved in [2], where it was shown that the conjecture would be true if the set is $\{1,2,\ldots,16\}$, and recently, it is shown in [10] that the conjecture would hold if the set is $\{1,2,3,4,5\}$.

Bartnicki, Grytczuk and Niwczyk [7] considered the choosability version of edge weighting. A graph is said to be *k-edge weight choosable* if the following is true: For any list assignment L which assigns to each edge e a set $L(e)$ of k real numbers, G has a proper edge weighting f such that $f(e) \in L(e)$ for each edge e. They conjectured that every graph without isolated edges is 3-edge weight choosable, and verified the conjecture for complete graphs, complete bipartite graphs and some other graphs.

A *total weighting* of G is a mapping $f : V \cup E \to \mathbb{R}$ which assigns to each vertex and each edge a real number as its weight. For a total weighting f, let $\varphi_f : V \to \mathbb{R}$ be defined as $\varphi_f(v) = \sum_{y \in E^*(v)} f(y)$. A total weighting is proper if the induced vertex colouring φ_f is proper. Przybyło and Woźniak [11, 12] studied total weighting of graphs. They conjectured that every simple graph G has a proper total weighting f such that $f(y) \in \{1,2\}$ for all $y \in V \cup E$, and verified this conjecture for some special graphs, including complete graphs, 4-regular graphs and graphs G with $\chi(G) \leq 3$. They also proved that every simple graph G has a proper total weighting f such that $f(y) \in \{1,2,\ldots,11\}$ for all $y \in V \cup E$. This result was improved in [10] where it was shown f can be chosen so that $f(v) \in \{1,2\}$ for every vertex v and $f(e) \in \{1,2,3\}$ for every edge e.

The choosability version of total weighting was studied in [13]. For positive integers k, k', a (k,k')-total list assignment is a mapping L which assigns to each vertex v a set $L(v)$ of k real numbers as permissible weights, and assigns to each edge e a set $L(e)$ is k' real numbers as permissible weights. A graph is called (k,k')-total weight choosable if for any (k,k')-total list assignment L, G has a proper total weighting f such that $f(y) \in L(y)$ for all $y \in V \cup E$. In the following, we say a graph is (k,k')-choosable if it is (k,k')-total weight choosable. It is known [13] that a graph is $(k,1)$-choosable if and only if it is k-choosable. So the concept of (k,k')-choosable builds a bridge between the concept of vertex colouring and edge weighting. The following conjectures were proposed in [13]:

Conjecture 1.1. Every graph is $(2,2)$-choosable.

Conjecture 1.2. Every graph with no isolated edges is $(1,3)$-choosable.

However, it is still unknown if there are constants k,k' such that every graph is (k,k')-choosable. It was shown in [13] that complete graphs, trees,

cycles, generalized theta graphs are $(2,2)$-choosable, and complete bipartite graphs $K_{2,n}$ are $(1,2)$-choosable and $K_{3,n}$ are $(2,2)$-choosable.

The method used in [7] and in [13] for the study of edge weight choosability or total weight choosability of graphs is algebraic (see Section 5), which proves the existence of a proper edge weighting or proper total weighting without actually constructing such a weighting.

In this paper, we introduce a method that constructs explicitly a proper total weighting of given graphs. Then we apply this method and prove that complete multipartite graphs of the form $K_{n,m,1,1,\ldots,1}$ are $(2,2)$-choosable and complete bipartite graphs other than K_2 are $(1,2)$-choosable.

2. The Max-Min Weighting Method

Suppose L is a (k,k')-total list assignment of a graph G. The max-min weighting method finds a proper L-weighting of G step by step. At each step, we find a vertex v and determine the weights of all the elements of $E^*(v)$. Let $f_0 : E^*(v) \to \mathbb{R}$ be a map such that for each $y \in E^*(v)$, $f_0(y) \in L(y)$. We call f_0 a *partial L-weighting of G for vertex v*. If f_0 is a partial L-weighting of G for v, let $\varphi_{f_0}(v) = \sum_{y \in E^*(v)} f_0(y)$. Then for any total weighting f of G which coincides with f_0 on $E^*(v)$, we have $\varphi_f(v) = \varphi_{f_0}(v)$. I.e., the colour of v is determined by the partial L-weighting f_0.

Given a partial L-weighting f_0 of G for vertex v. We define a (k,k')-total list assignment L_1 for $G_1 = G - v$ as follows:

$$L_1(y) = \begin{cases} L(y), & \text{if } y \text{ is a vertex not adjacent to } v \text{ or } y \text{ is an edge of } G_1, \\ \{w + f_0(e) : w \in L(y)\}, & \\ & \text{if } y \text{ is a vertex adjacent to } v \text{ and } e = yv. \end{cases}$$

We call L_1 *the (k,k')-total list assignment induced by L and f_0*.

Suppose L_1 is the (k,k')-total list assignment of G_1 induced by L and f_0. If f_1 is an L_1-weighting of G_1, then we combine f_1 and f_0 to obtain a

weighting f of G as follows:

$$f(y) = \begin{cases} f_1(y) - f_0(e), \\ \qquad \text{if } y \text{ is a vertex adjacent to } v \text{ and } e = yv, \\ f_1(y), \quad \text{if } y \text{ is a vertex not adjacent to } v \text{ or } y \text{ is an edge of } G_1, \\ f_0(y), \quad \text{if } y \in E^*(v) \end{cases}$$

We call f the L-weighting of G *induced by* f_1 *and* f_0. The following lemma is obvious and its proof is omitted.

Lemma 2.1. *Suppose L is a (k,k')-total list assignment of a graph G, f_0 is a partial L-weighting of G for vertex v, and L_1 is the (k,k')-total list assignment of $G_1 = G - v$ induced by L and f_0. If f_1 is an L_1-weighting of G_1, and f is the L-weighting of G induced by f_1 and f_0, then for each $u \in V(G_1)$, $\varphi_f(u) = \varphi_{f_1}(u)$ and $\varphi_f(v) = \varphi_{f_0}(v)$.* ■

For each $y \in V \cup E$, let $L_{\max}(y) = \max L(y)$ be the maximum number in $L(y)$, and let $L_{\min}(y) = \min L(y)$ be the minimum number in $L(y)$. Note that if $\big|L(y)\big| = 1$, then $L_{\max}(y) = L_{\min}(y)$. Otherwise, $L_{\max}(y) > L_{\min}(y)$.

Let

$$\varphi_{L,\max}(v) = \sum_{y \in E^*(v)} L_{\max}(y)$$

$$\varphi_{L,\min}(v) = \sum_{y \in E^*(v)} L_{\min}(y).$$

It follows from the definition that $\varphi_{L,\max}(v) = \max \varphi_f(v)$ and $\varphi_{L,\min}(v) = \min \varphi_f(v)$, where the maximum and minimum is taken over all L-weightings f of G. So for any L-weighting f of G, we have

$$\varphi_{L,\min}(v) \le \varphi_f(v) \le \varphi_{L,\max}(v).$$

For a vertex v of G, $N_G(v) = \{u : u \text{ is a neighbour of } v\}$, and $N_G[v] = N_G(v) \cup \{v\}$. A vertex v of G is *locally maximum (with respect to L)* if for any $u \in N_G(v)$, $\varphi_{L,\max}(u) \le \varphi_{L,\max}(v)$, and v is *locally minimum (with respect to L)* if for any $u \in N_G(v)$, $\varphi_{L,\min}(u) \ge \varphi_{L,\min}(v)$.

A partial L-weighting of G for a vertex v is said to *max-weights* v if $f(y) = L_{\max}(y)$ for each $y \in E^*(v)$. A partial L-weighting of G for a vertex

v is said to *min-weights* v if $f(y) = L_{\min}(y)$ to for each $y \in E^*(v)$. If f is a partial L-weighting of G that max-weights v, then $\varphi_f(v) = \varphi_{L,\max}(v)$; if f is a partial weighting of G that min-weights v, then $\varphi_f(v) = \varphi_{L,\min}(v)$.

Suppose we have defined a partial L-weighting f_0 of G for vertex v. Let L_1 be the (k,k')-total list assignment induced by L and f_0. By Lemma 2.1, to find a proper L-weighting of G, it suffices to find a proper L_1-weighting f_1 of G_1 so that for each $u \in N_G(v)$, $\varphi_{f_1}(u) \neq \varphi_{f_0}(v)$.

Our max-min-weighting method will only max-weight locally maximum vertices, and min-weight locally minimum vertices. By doing so, the conflicts with vertices "coloured" in later steps will not be too complicated. Indeed, assume v is a locally maximum vertex, and f_0 is a partial L-weighting of G for v that max-weights v. Let

$$Q_{L,\max}(v) = \left\{ u \in N_G(v) \,:\, \varphi_{L,\max}(u) = \varphi_{L,\max}(v) \right\}.$$

To extend f_0 to a proper L-weighting f of G, we need to make sure that no vertex $u \in Q_{L,\max}(v)$ will be "coloured" with $\varphi_{L,\max}(u)$. For this purpose, it suffices that some $y \in E^*(u)$ is assigned a weight smaller than $L_{\max}(y)$. I.e., u is not max-weighted (under the list L). Similarly, assume v is a locally minimum vertex, and f_0 is a partial L-weighting of G for v that min-weights v, and

$$Q_{L,\min}(v) = \left\{ u \in N_G(v) \,:\, \varphi_{L,\min}(u) = \varphi_{L,\min}(v) \right\}.$$

To extend f_0 to a proper L-weighting f of G, we need to make sure that $u \in Q_{L,\min}(v)$ is not min-weighted (under the list L). This motivates the following definition.

Definition 2.2. Suppose $G = (V,E)$ is a graph and L is a (k,k')-total list assignment of G. Let S_0, T_0 be two subsets of V (each of S_0, T_0 may be empty and S_0, T_0 need not be disjoint). A *proper* $(L; S_0, T_0)$*-weighting* of G is a proper L-weighting f of G such that for each $v \in S_0$, $\varphi_f(v) \neq \varphi_{L,\max}(v)$ and for each $v \in T_0$, $\varphi_f(v) \neq \varphi_{L,\min}(v)$.

The following theorem contains the key idea of max-min weighting method.

Theorem 2.3. *Suppose $G = (V,E)$ is a graph, $k' \geq 2$ and L is a (k,k')-total list assignment of G, and S_0, T_0 are two subsets of V.*

- *Assume v is a locally maximum vertex and $v \notin S_0$. In case $k = 1$, we further assume that v is not an isolated vertex. Let f_0 be the partial*

L-weighting of G that max-weights v. Let $G_1 = G - v$, and let L_1 be the (k, k')-total list assignment induced by L and f_0. Let

$$S_1 = S_0 \cup Q_{L,\max}(v), \quad T_1 = T_0 - N_G[v].$$

If f_1 is a proper $(L_1; S_1, T_1)$-weighting of G_1, then the L-weighting induced by f_1 and f_0 is a proper $(L; S_0, T_0)$-weighting of G.

- *Assume v is a locally minimum vertex and $v \notin T_0$. In case $k = 1$, we further assume that v is not an isolated vertex. Let f_0 be the partial L-weighting of G that min-weights v. Let $G_1 = G - v$, and let L_1 be the (k, k')-total list assignment induced by L and f_0. Let*

$$S_1 = S_0 - N_G[v], \quad T_1 = T_0 \cup Q_{L,\min}(v).$$

If f_1 is a proper $(L_1; S_1, T_1)$-weighting of G_1, then the L-weighting induced by f_1 and f_0 is a proper $(L; S_0, T_0)$-weighting of G.

Proof. Assume v is a locally maximum vertex and $v \notin S_0$. Let f_0 be the partial L-weighting of G that max-weights v. Let G_1, S_1, T_1 be defined as above. Assume f_1 is a proper $(L_1; S_1, T_1)$-weighting of G_1. Let g be the L-weighting of G induced by f_1 and f_0. To prove that g is a proper $(L; S_0, T_0)$-weighting of G, it suffices to show the following:

1. For any $u \in Q_{L,\max}(v)$, $\varphi_g(u) \neq \varphi_{L,\max}(u)$.
2. For any $u \in S_0$, $\varphi_g(u) \neq \varphi_{L,\max}(u)$.
3. For any $u \in T_0$, $\varphi_g(u) \neq \varphi_{L,\min}(u)$.

By Lemma 2.1, $\varphi_g(u) = \varphi_{f_1}(u)$ for $u \neq v$ and $\varphi_g(v) = \varphi_{L,\max}(v)$. Since $Q_{L,\max}(v) \subseteq S_1$ and f_1 is a proper $(L_1; S_1, T_1)$-weighting of G, it follows that for $u \in Q_{L,\max}(v) \subseteq S_1$, $\varphi_g(u) = \varphi_{f_1}(u) \neq \varphi_{L,\max}(u)$. So (1) is satisfied. Since $S_0 \subseteq S_1$, we also have $\varphi_g(u) = \varphi_{f_1}(u) \neq \varphi_{L,\max}(u)$ for every $u \in S_0$. So (2) is satisfied. Assume $u \in T_0$. If $u \in T_1$, then $\varphi_g(u) = \varphi_{f_1}(u) > \varphi_{L_1,\min}(u) = \varphi_{L,\min}(u)$. If $u \in N_G(v)$, then since the edge $e = uv$ is assigned the weight $g(e) = f_0(e) = L_{\max}(e) > L_{\min}(e)$ (here we use the assumption that $k' \geq 2$), we conclude that $\varphi_{L_1,\min}(u) > \varphi_{L,\min}(u)$. Therefore $\varphi_g(u) \geq \varphi_{L_1,\min}(u) > \varphi_{L,\min}(u)$. If $u = v$, then since either $k \geq 2$ or v is not isolated, we have $\varphi_{L,\max}(v) > \varphi_{L,\min}(v)$ (here again we use the assumption that $k' \geq 2$). Therefore $\varphi_g(v) = \varphi_{L,\max}(v) \neq \varphi_{L,\min}(v)$. Thus (3) is satisfied.

The case that v is locally minimum is symmetric, and the proof is the same, except that max is replaced by min, $Q_{L,\max}(v)$ is replaced by $Q_{L,\min}(v)$, and S_0, T_0 are interchanged, and S_1, T_1 are interchanged. ■

3. Application of the Max-Min Method

In this section, we use the max-min weighting method to prove that complete multipartite graphs of the form $K_{n,m,1,1,\ldots,1}$ are $(2,2)$-choosable and complete bipartite graphs other than K_2 are $(1,2)$-choosable.

In most of the applications of the max-min weighting method in this paper, we max-weight globally maximum vertices and min-weight globally minimum vertices. For a (k,k')-total list assignment L of G, let $V_{L,\max}(G)$ be the set of globally maximum vertices of G and let $V_{L,\min}(G)$ be the set of globally minimum vertices of G. To be precise,

$$V_{L,\max}(G) = \{v \in V(G) : \forall u \in V(G),\ \varphi_{L,\max}(v) \geq \varphi_{L,\max}(u)\}$$

$$V_{L,\min}(G) = \{v \in V(G) : \forall u \in V(G),\ \varphi_{L,\min}(v) \leq \varphi_{L,\min}(u)\}.$$

First we prove that complete multipartite graphs of the form $K_{n,m,1,1,\ldots,1}$ are $(2,2)$-choosable.

Theorem 3.1. *Suppose $n, m, k \geq 0$ are integers. Let $G = K_{n,m,1,1,\ldots,1}$ be the complete multipartite graph with one partite set of size n, one partite set of size m and k partite sets of size 1. Denote by A the partite set of size n, and by B the partite set of size m and by C the union of the k partite sets each of size 1. Suppose L is a $(2,2)$-total list assignment L of G, and S_0, T_0 are subsets of $V(G)$ such that one of S_0, T_0 is empty and the other has an empty intersection with A or B. Then there is a proper $(L; S_0, T_0)$-weighting of G. Hence G is $(2,2)$-choosable.*

Proof. We prove this theorem by induction on $|V(G)|$. If $|V(G)| = 1$, then the conclusion is obviously true. Assume $|V(G)| > 1$. Without loss of generality, we assume that $S_0 = \emptyset$ and assume that $T_0 \cap A = \emptyset$.

- If $V_{L,\max}(G) \cap A \neq \emptyset$, then let $v \in V_{L,\max}(G) \cap A$.
- If $V_{L,\max}(G) \cap A = \emptyset$ and $V_{L,\max}(G) \cap C \neq \emptyset$, then let $v \in V_{L,\max}(G) \cap C$.
- If $V_{L,\max}(G) \cap (A \cup C) = \emptyset$, then let $v \in V_{L,\max}(G) \cap B$.

Let f_0 be the partial L-weighting of G that max-weights v. Let $G' = G - v$ and L_1 be the $(2,2)$-total list assignment induced by L and f_0. Let $S_1 = S_0 \cup Q_{L,\max}(v) = Q_{L,\max}(v)$ and let $T_1 = T_0 - N_G[v]$. It is easy to verify that if $v \in A$, then $T_1 = \emptyset$ and $S_1 \cap A = \emptyset$. If $v \in C$, then $T_1 = \emptyset$ and

$S_1 \cap A = \emptyset$. If $v \in B$, then $S_1 = \emptyset$ and $T_1 \cap A = \emptyset$. By induction hypothesis, G' has a proper $(L_1; S_1, T_1)$-weighting. By Theorem 2.3, G has a proper $(L; S_0, T_0)$-weighting. ■

It was proved in [13] that complete graphs are $(2,2)$-choosable. The proof in [13] uses Combinatorial Nullstellensatz and does not provide an algorithm that constructs a proper L-total weighting of K_n for a given $(2,2)$-total list assignment L of K_n. The proof above gives a linear time algorithm that constructs a required proper L-total weighting of $K_{n,m,1,1,\ldots,1}$.

In [13], the problem whether complete bipartite graphs are $(2,2)$-choosable was studied. It was proved there that for any positive integer n, $K_{2,n}$ are $(1,2)$-choosable and $K_{3,n}$ are $(2,2)$-choosable. But the question whether all complete bipartite graphs are $(2,2)$-choosable remained open. Theorem 3.1 answers this question in the affirmative. Theorem 3.2 below shows that complete bipartite graphs are actually $(1,2)$-choosable, provided that it is not K_2.

Theorem 3.2. *Suppose $G = K_{n,m} \neq K_2$ is a complete bipartite graph and L is a $(1,2)$-total list assignment of G. If both S_0, T_0 are empty sets, then G has a proper $(L; S_0, T_0)$-weighting.*

Proof. Assume the theorem is not true and $G = K_{n,m}$ is a counterexample with minimum number of vertices. Let A, B be the two partite sets with $|A| = n$ and $|B| = m$.

Claim 1. $G \notin \{K_{1,2}, K_{1,3}, K_{2,2}, K_{2,3}\}$.

The graphs excluded in this claim are small graphs. One way to prove this claim is to do a case by case check, which seems to be tedious. Another method is to use Combinatorial Nullstellensatz, which will be discussed in Section 5.

Claim 2. None of $V_{L,\max}(G)$ and $V_{L,\min}(G)$ is an independent set of G.

Proof. Assume to the contrary that $V_{L,\max}(G)$ is an independent set of G, i.e., $V_{L,\max}(G)$ is a subset of A.

Let $v \in V_{L,\max}(G)$ and let f_0 be the partial L-weighting of G that max-weights v. Let L_1 be the $(1,2)$-total list assignment of $G_1 = G - v$ induced by L and f_0. Since $S_0 = T_0 = \emptyset$, and $Q_{L,\max}(v) = \emptyset$, we have $S_1 = T_1 = \emptyset$ for the sets S_1, T_1 defined in Theorem 2.3. Since $G_1 \neq K_2$, by induction hypothesis, G_1 has a proper $(L_1; S_1, T_1)$-weighting f_1. By Theorem 2.3, the L-weighting f of G induced by f_1 and f_0 is a proper L-weighting of G. ■

Let

$$A_{\max} = V_{L,\max}(G) \cap A,$$

$$B_{\max} = V_{L,\max}(G) \cap B,$$

$$A_{\min} = V_{L,\min}(G) \cap A,$$

$$B_{\min} = V_{L,\min}(G) \cap B.$$

By Claim 2, none of the sets $A_{\max}, B_{\max}, A_{\min}, B_{\min}$ is empty.

Claim 3. $|A_{\max} \cup A_{\min}| = |B_{\max} \cup B_{\min}| = 1$.

Proof. Assume to the contrary that $|A_{\max} \cup A_{\min}| > 1$. Then there exist $v \in A_{\max}$ and $v' \in A_{\min}$ such that $v \neq v'$.

Let f_0 be the partial L-weighting of G that max-weights v, and let L_1 be the $(1,2)$-total weighting of $G_1 = G - v$ induced by L and f_0. Then for any vertex $x \in B$,

$$\varphi_{L_1,\min}(x) > \varphi_{L,\min}(x) \geq \varphi_{L,\min}(v') = \varphi_{L_1,\min}(v').$$

Therefore $v' \in V_{L_1,\min}(G_1)$ and $V_{L_1,\min}(G_1) \cap B = \emptyset$.

Let $S_1 = S_0 \cup Q_{L,\max}(v)$ and $T_1 = T_0 - N_G[v]$. We have $T_1 = \emptyset$ and $S_1 = B_{\max}$.

Let f_1 be the partial L_1-weighting of G_1 that min-weights v' and let L_2 be the $(1,2)$-total list assignment of $G_2 = G_1 - v'$ induced by L_1 and f_1. Let $S_2 = S_1 - N_{G_1}[v']$ and $T_2 = T_1 \cup Q_{L_1,\min}(v')$. It is obvious that $S_2 = \emptyset$. Since $V_{L_1,\min}(G_1) \cap B = \emptyset$, we have $T_2 = Q_{L_1,\min}(v') = \emptyset$. By the minimality of G, we know that G_2 has a proper $(L_2; S_2, T_2)$-weighting. By Theorem 2.3, G_1 has a proper $(L_1; S_1, T_1)$-weighting and G has a proper $(L; S_0, T_0)$-weighting. ■

Assume $A_{\max} = A_{\min} = \{v\}$ and $B_{\max} = B_{\min} = \{u\}$.

Let f_0 be the partial L-weighting of G that max-weights v, and let L_1 be the $(1,2)$-total list assignment of $G_1 = G - v$ induced by L and f_0. Then $S_1 = S_0 \cup Q_{L,\max}(v) = \{u\}$ and $T_1 = T_0 - N_G[v] = \emptyset$. By Theorem 2.3, if G_1 has a proper $(L_1; S_1, T_1)$-weighting, then G has a proper $(L; S_0, T_0)$-weighting. As $S_1 \neq \emptyset$, we cannot conclude (from the minimality of G) that G_1 has a proper $(L_1; S_1, T_1)$-weighting.

If $V_{L_1,\min}(G_1) \subseteq A$, then let $v' \in V_{L_1,\min}(G_1)$. Let f_1 be the partial L_1-weighting that min-weights v'. Let L_2 be the $(1,2)$-total list assignment

of $G_2 = G_1 - v'$ induced by L_1 and f_1. Let $S_2 = S_1 - N_G[v']$ and $T_2 = T_1 \cup Q_{L_1,\min}(v')$. Then $S_2 = T_2 = \emptyset$. Since $G_2 \neq K_2$, by the minimality of G, G_2 has a proper $(L_2; S_2, T_2)$-weighting. By Theorem 2.3, this implies that G_1 has a proper $(L_1; S_1, T_1)$-weighting, and hence G has a proper $(L; S_0, T_0)$-weighting.

Thus we may assume that $V_{L_1,\min}(G_1) \cap B \neq \emptyset$. Let

$$B' = V_{L_1,\min}(G_1) \cap B.$$

By definition, for each $x \in B'$, for each $y \in A - \{v\}$, $\varphi_{L_1,\min}(x) \leq \varphi_{L_1,\min}(y)$. However, it follows from the definition that $\varphi_{L_1,\min}(x) > \varphi_{L,\min}(x)$ and $\varphi_{L_1,\min}(y) = \varphi_{L,\min}(y)$. So the following claim holds.

Claim 4. For each $x \in B'$ and for each $y \in A - \{v\}$, $\varphi_{L,\min}(x) < \varphi_{L,\min}(y)$.

Interchange the roles of A and B, and let f_0' be the partial L-weighting of G that max-weights u, let L_1' be the $(1,2)$-weighting of $G_1' = G - u$ induced by L and f_0', and let

$$A' = V_{L_1',\min}(G_1') \cap A.$$

By symmetry, we have $A' \neq \emptyset$ and the following claim holds.

Claim 5. For each $y \in A'$ and for each $x \in B - \{u\}$, $\varphi_{L,\min}(y) < \varphi_{L,\min}(x)$.

By combining Claims 4 and 5, we can prove the following claim.

Claim 6. $A' = \{v\}$ and $B' = \{u\}$.

Proof. If $A' - \{v\} \neq \emptyset$ and $B' - \{u\} \neq \emptyset$, then for $y \in A' - \{v\}$ and $x \in B' - \{u\}$, Claims 4 and 5 contradicts each other. Thus, by symmetry, we may assume that $B' - \{u\} = \emptyset$ and hence $B' = \{u\}$ (as $B' \neq \emptyset$).

Assume the claim is not true. Then $A' - \{v\} \neq \emptyset$. Let $s \in A' - \{v\}$. By definition, $\varphi_{L,\min}(s) < \varphi_{L_1',\min}(s) \leq \varphi_{L_1',\min}(v)$. Let $e = uv$. Then

$$\varphi_{L_1,\min}(u) = \varphi_{L,\min}(u) + L_{\max}(e) - L_{\min}(e)$$

$$\varphi_{L_1',\min}(v) = \varphi_{L,\min}(v) + L_{\max}(e) - L_{\min}(e).$$

As $\varphi_{L,\min}(u) = \varphi_{L,\min}(v)$, we have $\varphi_{L_1,\min}(u) = \varphi_{L_1',\min}(v)$. Hence

$$\varphi_{L_1,\min}(s) = \varphi_{L,\min}(s) < \varphi_{L_1,\min}(u).$$

But $B' = \{u\}$ implies that $\varphi_{L_1,\min}(u) \leq \varphi_{L_1,\min}(x)$ for all $x \in A - \{v\}$. This is a contradiction. ■

Let $w \neq u, v$ be a vertex such that for any $x \neq u, v$, $\varphi_{L,\max}(x) \leq \varphi_{L,\max}(w)$. Without loss of generality, we assume that $w \in B$. Then in the graph $G_1 = G - v$, w is a local maximum vertex with respect to L_1. Let f_1 be the partial L_1-weighting of G_1 that max-weights w. Let L_2 be the $(1,2)$-total list assignment of $G_2 = G_1 - w$ induced by L_1 and f_1. Let $S_2 = S_1 \cup Q_{L_1,\max}(w)$ and $T_2 = T_1 - N_{G_1}[w]$. Then $T_2 = \emptyset$ and $S_2 \subseteq \{u\} \cup \big(A - \{v\}\big)$.

Now we shall prove that $V_{L_2,\min}(G_2) = \{u\}$. I.e., for any $z \in \big(A - \{v\}\big) \cup \big(B - \{u, w\}\big)$,

$$\varphi_{L_2,\min}(u) < \varphi_{L_2,\min}(z).$$

If $z \in B - \{u, w\}$, then since $B' = \{u\}$, it follows that

$$\varphi_{L_2,\min}(z) = \varphi_{L_1,\min}(z) > \varphi_{L_1,\min}(u) = \varphi_{L_2,\min}(u).$$

If $z \in A - \{v\}$, then

$$\varphi_{L_2,\min}(z) > \varphi_{L_1,\min}(z) \geq \varphi_{L_1,\min}(u) = \varphi_{L_2,\min}(u).$$

Let f_2 be the partial L_2-weighting of G_2 that min-weights u, and let L_3 be the $(1,2)$-total list assignment of $G_3 = G_2 - u$ induced by L_2 and f_2. Let $S_3 = S_2 - N_{G_2}[u]$ and $T_3 = T_2 \cup Q_{L_2,\min}(u)$. Then $S_3 = T_3 = \emptyset$. Since $G \neq K_{2,3}$, hence $G_3 \neq K_2$. Hence G_3 has a proper $(L_3; S_3, T_3)$-weighting. By repeatedly applying Theorem 2.3, we conclude that G_2 has a proper $(L_2; S_2, T_2)$-weighting, and hence G_1 has a proper $(L_1; S_1, T_1)$-weighting, and hence G has a proper $(L; S_0, T_0)$-weighting. ■

It was shown in [7] that complete bipartite graphs other than K_2 are 3-edge weight choosable. By assigning to every vertex weight 0, we have the following corollary.

Corollary 3.3. *Every complete bipartite graphs other than K_2 are 2-edge weight choosable.*

Corollary 3.3 can be stated as a result about matrices, which seems interesting.

Corollary 3.4. *If $n + m \geq 3$, and $\mathcal{L} = [L_{ij}]$ is an $n \times m$ matrix where each L_{ij} is a set of two real numbers, then there is an $n \times m$ real matrix $A = [a_{ij}]$ such that $a_{ij} \in L_{ij}$ and no row sum of A is equal to a column sum of A.*

Note that the induction proof of Theorem 3.2 cannot be used to give a direct proof of Corollary 3.3. This shows that even if our interests are in edge weight choosability of graphs, it can be helpful by considering the more general total weight choosability of graphs.

4. Algebraic Method

Assume $G = (V, E)$ is a simple graph, where $V = \{v_1, v_2, \ldots, v_n\}$ and $E = \{e_1, e_2, \ldots, e_m\}$. Denote by X the set of variables $x_{v_1}, x_{v_2}, \ldots, x_{v_n}, x_{e_1}, x_{e_2}, \ldots, x_{e_m}$. For each vertex v of G, let $Q_v = \sum_{e \in E(v)} x_e + x_v$. Fix an arbitrary orientation $\vec{G}$ of G. Let

$$\tilde{P}(\vec{G}) = \tilde{P}(x_{v_1}, x_{v_2}, \ldots, x_{v_n}, x_{e_1}, x_{e_2}, \ldots, x_{e_m}) = \prod_{uv \in E(\vec{G})} (Q_v - Q_u).$$

For each edge uv (oriented from u to v), $Q_v - Q_u$ is a polynomial of degree 1 with variable set X. So $\tilde{P}(\vec{G})$ is a polynomial of degree $m = \big|E(G)\big|$ with variable set X. For different orientations $\vec{G}$ of G, the polynomials defined may differ by a sign, which is irrelevant for our purpose.

Suppose f is a total weighting of G. By definition, for each vertex u,

$$\varphi_f(u) = Q_u\big(f(v_1), f(v_2), \ldots, f(v_n), f(e_1), f(e_2), \ldots, f(e_m)\big)$$

is equal to the evaluation of the polynomial Q_u with $x_v = f(v)$ for each $v \in V$ and $x_e = f(e)$ for each edge $e \in E$. Therefore f is a proper total weighting of G if and only if

$$\tilde{P}\big(f(v_1), f(v_2), \ldots, f(v_n), f(e_1), f(e_2), \ldots, f(e_m)\big) \neq 0.$$

A mapping $\eta : V \cup E \to \{0, 1, \ldots\}$ is called a *valid index function* if $\sum_{y \in V \cup E} \eta(y) = m$. For a valid index function η, let c_η be the coefficient of the monomial $\prod_{y \in V \cup E} x_y^{\eta(y)}$ in the expansion of $\tilde{P}(\vec{G})$. It follows from the Combinatorial Nullstellensatz [6, 5] that if $c_\eta \neq 0$, and $\big|L(y)\big| \geq \eta(y) + 1$ for all $y \in V \cup E$, then there exist $f(y) \in L(y)$ for $y \in V \cup E$ such that

$$\tilde{P}\big(f(v_1), f(v_2), \ldots, f(v_n), f(e_1), f(e_2), \ldots, f(e_m)\big) \neq 0.$$

I.e., G has a proper L-total weighting.

The coefficient c_η of the monomial $\prod_{y\in V\cup E} x_y^{\eta(y)}$ is associated to the permanent of some matrices. Let $A_{\vec{G}}$ be the $m \times (n+m)$ matrix, with rows indexed by the edges of $\vec{G}$ and columns indexed by vertices and edges, defined as follows: If $e = uv$ is an edge of $\vec{G}$ oriented from u to v, then let

$$a_{ey} = \begin{cases} 1, & \text{if } y = v \text{ or } y \neq e \text{ is an edge incident to } v \\ -1, & \text{if } y = u \text{ or } y \neq e \text{ is an edge incident to } u \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$\tilde{P}(\vec{G}) = \prod_{e\in E} \Big(\sum_{y\in V\cup E} a_{ey} x_y \Big).$$

For each $y \in V \cup E$, we denote by A_y the column of $A_{\vec{G}}$ indexed by y. So

$$A_{\vec{G}} = [A_{v_1}, A_{v_2}, \ldots, A_{v_n}, A_{e_1}, A_{e_2}, \ldots, A_{e_m}].$$

For a valid index function η, let $A_{\vec{G}}(\eta)$ be the $m \times m$ matrix in which the column A_y occurs $\eta(y)$ times for each $y \in V \cup E$. It is known [4] that $c_\eta \neq 0$ if and only if the permanent of $A_{\vec{G}}(\eta)$ is non-zero, where the permanent of an $m \times m$ matrix $A = [a_{ij}]$ is defined as

$$\mathrm{per}\,(A) = \sum_{\sigma} \prod_{j=1}^{m} a_{j\sigma(j)},$$

here the summation is taken over all the permutations σ of $\{1, 2, \ldots, m\}$.

Thus by showing that certain matrices have non-zero permanent, we can prove results about total weight choosability of graphs. For any graph $G = (V, E)$, let η^*_G be the valid index function defined as $\eta^*_G(e) = 1$ for each edge e and $\eta^*_G(v) = 0$ for each vertex v. By the discussion above, if $A_{\vec{G}}(\eta^*_G)$ has non-zero permanent, then G is $(1,2)$-choosable. It is not difficult to verify that for $G = K_{1,2}, K_{1,3}, K_{2,2}, K_{2,3}$, the permanent of $A_{\vec{G}}(\eta^*_G)$ is non-zero. Therefore these graphs are $(1,2)$-choosable.

We believe that Conjectures 1.1 and 1.2 can be proved by using this method. To be precise, the following stronger conjectures are proposed in [7, 13]:

Conjecture 4.1 [13]. For any graph G, there is a valid index function η with $\eta(y) \leq 1$ for all $y \in V \cup E$ such that $\mathrm{per}\,(A_{\vec{G}}(\eta)) \neq 0$.

Conjecture 4.2 [7]. For any graph G without isolated edges, there is a valid index function η with $\eta(v) = 0$ for all $v \in V$ and $\eta(e) \leq 2$ for all $e \in E$ such that per $\left(A_{\vec{G}}(\eta)\right) \neq 0$.

In [13], we had limited success and proved that if $G = K_{2,n}$, then G is $(1,2)$-choosable, and if G is a complete graph, or $G = K_{3,n}$, or G is a tree or a generalized theta graph, then Conjecture 4.1 holds for G, and hence G is $(2,2)$-choosable. We were unable to prove that general complete bipartite graphs are $(2,2)$-choosable by this method in [13]. In this paper, by using the max-min weighting method, we are able to prove a stronger result: if G is a complete bipartite graph, then G is $(1,2)$-choosable. The main result in [13] that complete graphs are $(2,2)$-choosable is also implied by Theorem 3.1. In this sense, the max-min method has been more successful. However, it seems to us that the hope of confirming Conjectures 1.1 and 1.2 is more likely to lie on the algebraic method. For this purpose, we would still like to see a proof of the results in this paper by the algebraic method. Because complete bipartite graphs other than K_2 are $(1,2)$-choosability, one naturally asks whether per $\left(A_{\vec{G}}(\eta^*_G)\right) \neq 0$ for all complete bipartite graphs $G \neq K_2$. But this question has a negative answer.

Theorem 4.3. *If G is a bipartite graph with an odd number of edges and there is an automorphism of G that interchanges the two partite sets, then* per $\left(A_{\vec{G}}(\eta^*_G)\right) = 0$. *In particular, if n is odd and $G = K_{n,n}$, then* per $\left(A_{\vec{G}}(\eta^*_G)\right) = 0$.

Proof. The permanent of the matrix $A_{\vec{G}}(\eta^*_G)$ depends on the orientation of the edges. Changing the orientation of one edge will change the sign of the permanent. Assume the two partite sets of G are A and B. Let $\vec{G}_1$ be the orientation of G in such a way that all the edges are from B to A. Let $\vec{G}_2$ be the orientation of G in which all the edges are oriented from A to B. Since $\vec{G}_2$ is obtained from $\vec{G}_1$ by changing the orientation of an odd number of edges, their permanents differ by a sign. However, by our assumption, $\vec{G}_2$ and $\vec{G}_1$ are isomorphic, so they have the same permaments. Therefore their common permanent is 0. ■

This result shows that not every result on total weight choosability of graphs can be proved by the method of calculating the permanents of some matrices. Nevertheless, it has not yet crushed our faith on Conjectures 4.1 and 4.2. For complete bipartite graphs, based on some experiments, we suspect that if $n \neq m$ or $n = m$ are even, then for $G = K_{n,m}$, $A_{\vec{G}}(\eta^*_G)$ has nonzero permanent. We can prove this for some special cases.

Theorem 4.4. *If n, q are positive integers such that n has a prime factor $p > q$, then for $G = K_{n,q}$, $A_{\vec{G}}(\eta_G^*)$ has nonzero permanent.*

Proof. As the implication of this theorem to total choosability of graphs is already proved, we shall just sketch an outline of the proof and omit some details.

Assume the two partite sets of G are $A = \{v_1, v_2, \ldots, v_n\}$ and $B = \{u_1, u_2, \ldots, u_q\}$. We orient the edges of G so that all the edges are oriented from B to A. We shall prove that $\operatorname{per}\left(A_{\vec{G}}(\eta_G^*)\right)$ is not a multiple of p, and hence $\operatorname{per}\left(A_{\vec{G}}(\eta_G^*)\right) \neq 0$.

Suppose $e = uv$ is an edge of G oriented from u to v, then the column A_e can be expressed as $A_e = A_v + A_u$. It is known (and easy to see) that the following is true: Assume A is a matrix, and A_i is a column of A, and $A_i = A_i' + A_i''$. Let A' be obtained from A by replacing A_i with A_i' and A'' be obtained from A by replacing A_i by A_i''. Then

$$\operatorname{per}(A) = \operatorname{per}(A') + \operatorname{per}(A''). \tag{1}$$

The following observation can be derived from the definition of permanent (cf. [4])

Observation 4.5. *If M is an integer matrix, and k of the columns of M are identical, then* $\operatorname{per}(M)$ *is a multiple of $k!$.*

Suppose F is a subset of E and F' is a subset of F. Let $A_{F',F}$ be the matrix obtained from $A_{\vec{G}}(\eta_G^*)$ by replacing, for each $e = uv \in F'$, the column A_e by A_u, and replacing for each $e = uv \in F - F'$, the column A_e by A_v. For any subset F of E, by repeatedly applying (1), we have

$$\operatorname{per}\left(A_{\vec{G}}(\eta_G^*)\right) = \sum_{F' \subseteq F} \operatorname{per}(A_{F',F}).$$

Let $F_1 = \{u_1 v_j : j = 1, 2, \ldots, n\}$. For $k = 0, 1, \ldots, n$, let

$$\mathcal{F}_{1,k} = \left\{F' \subseteq F_1 : |F'| = k\right\}.$$

It is easy to see that if $F', F'' \in \mathcal{F}_{1,k}$, then

$$\operatorname{per}(A_{F',F_1}) = \operatorname{per}(A_{F'',F_1}).$$

If $|F'| = k$ and $k \geq p$, then by Observation 4.5, $\left|\operatorname{per}(A_{F',F_1})\right|$ is a multiple of $k!$, which is a multiple of p.

If $1 \leq k < p$, then $|\mathcal{F}_{1,k}| = \binom{n}{k}$ is a multiple of p. Hence the sum $\sum_{F' \in \mathcal{F}_{1,k}} \operatorname{per}(A_{F',F_1})$ is a multiple of p. Therefore

$$\operatorname{per}\left(A_{\vec{G}}(\eta_G^*)\right) \cong \operatorname{per}(A_{\emptyset,F_1}) \pmod{p}.$$

Let $F_2 = \{u_j v_s : s = 1, 2, \ldots, n,\ j = 1, 2\}$. Then

$$\operatorname{per}(A_{\emptyset,F_1}) = \sum_{F' \subseteq F_2,\ F' \cap F_1 = \emptyset} \operatorname{per}(A_{F',F_2}).$$

Repeat the above process, we can prove that

$$\operatorname{per}(A_{\emptyset,F_1}) \cong \operatorname{per}(A_{\emptyset,F_2}) \pmod{p}.$$

For $i \geq 2$, let $F_i = \{u_j v_s : s = 1, 2, \ldots, n,\ j = 1, 2, \ldots, i\}$. By repeating the process above, we can prove that

$$\operatorname{per}\left(A_{\vec{G}}(\eta_G^*)\right) \cong \operatorname{per}(A_{\emptyset,F_1}) \cong \cdots \cong \operatorname{per}(A_{\emptyset,F_i}) \pmod{p}.$$

Therefore

$$\operatorname{per}\left(A_{\vec{G}}(\eta_G^*)\right) \cong \operatorname{per}(A_{\emptyset,E}) \pmod{p}.$$

But it is easy to verify that $\operatorname{per}(A_{\emptyset,E}) = (q!)^n \ncong 0 \pmod{p}$. ■

It would be interesting to prove that for any complete bipartite graph $G = K_{n,m}$, unless $n = m$ is odd, we have $\operatorname{per}\left(A_{\vec{G}}(\eta_G^*)\right) \neq 0$. If this is true, then it follows from a result in [13] that Conjecture 4.1 holds for all complete bipartite graphs.

References

[1] L. Addario-Berry, R. E. L. Aldred, K. Dalal and B. A. Reed, Vertex colouring edge partitions, *J. Combin. Theory Ser. B,* **94** (2005), 237–244.

[2] L. Addario-Berry, K. Dalal and B. A. Reed, *Degree constrianed subgraphs,* Proceedings of GRACO2005, Volume 19, Electron. Notes Discrete Math., Amsterdam (2005), 257–263, Elsevier.

[3] L. Addario-Berry, K. Dalal, C. McDiarmid, B. A. Reed and A. Thomason, Vertex-colouring edge-weightings, *Combinatorica,* **27** (2007), 1–12.

[4] N. Alon and M. Tarsi, A nowhere zero point in linear mappings, *Combinatorica,* **9** (1989), 393–395.

[5] N. Alon and M. Tarsi, Colorings and orientations of graphs, *Combinatorica,* **12** (1992), 125–134.

[6] N. Alon and M. Tarsi, Combinatorial Nullstellensatz, *Combin. Prob. Comput.,* **8** (1999), 7–29.

[7] T. Bartnicki, J. Grytczuk and S. Niwczyk, Weight choosability of graphs, *Journal of Graph Theory,* **60(3)** (2009), 242–256.

[8] G. Chang, C. Lu, J. Wu and Q. Yu, *Vertex coloring 2-edge weighting of bipartite graphs,* preprint, 2007.

[9] M. Karoński, T. Łuczak and A. Thomason, Edge weights and vertex colour, *J. Combin. Theory Ser. B,* **91** (2004), 151–157.

[10] M. Kalkowski, M. Karoński and F. Pfender, Vertex-coloring edge-weightings: towards the 1-2-3-Conjecture, *Journal of Combinatorial Theory, Ser. B,* to appear.

[11] J. Przybyło and M. Woźniak, 1, 2-*conjecture,* Preprint MD 024, www.ii.uj.edu.pl/preMD/

[12] J. Przybyło and M. Woźniak, 1, 2-*conjecture II,* Preprint MD 026, www.ii.uj.edu.pl/preMD/

[13] T. Wong and X. Zhu, Total weight choosability of graphs, *Journal of Graph Theory,* to appear.

Tsai-Lien Wong
Department of Applied Mathematics
National Sun Yat-sen University
e-mail: tlwong@math.nsysu.edu.tw

Daqing Yang
Center for Discrete Mathematics
Fuzhou University
Fuzhou, Fujian 350002
China
e-mail: daqing85@yahoo.com

Xuding Zhu
Department of Applied Mathematics
National Sun Yat-sen University
Kaohsiung, Taiwan 80424
and
National Center for Theoretical Sciences
e-mail: zhu@math.nsysu.edu.tw

BOLYAI SOCIETY
MATHEMATICAL STUDIES, 20

Fete of Combinatorics, pp. 355–365.

Open Problems

1. The Grothendieck Constant and the Lovász Theta Function (An Open Problem)

Problem poser: Noga Alon (Department of Mathematics and Computer Science, Tel Aviv University, e-mail: nogaa@post.tau.ac.il)

The *Grothendieck constant* $K(G)$ of a graph $G = (V, E)$ is the least constant K such that for every $A : E \to R$,

$$\sup_{f:V\to S^{|V|-1}} \sum_{\{u,v\}\in E} A(u,v)\cdot\langle f(u), f(v)\rangle \leq K \sup_{f:V\to\{-1,+1\}} \sum_{\{u,v\}\in E} A(u,v)\cdot f(u)f(v).$$

This notion was introduced in [2], where it is shown that there is an absolute positive constant c so that for every graph G, $K(G) \leq c \log \vartheta(\overline{G})$, with $\vartheta(\overline{G})$ being the Lovász theta function of the complement of G, defined in [12].

Problem. Is there a lower bound for $K(G)$ in terms of $\vartheta(\overline{G})$? In particular, is there an absolute constant $c' > 0$ so that for every graph G, $K(G) \geq c' \log \vartheta(\overline{G})$?

The results in [2] imply that this is true for perfect graphs (and in fact for any family of graphs in which the chromatic number and the maximum clique are polynomially related), and the results in [1] show that this is true for random graphs and for certain pseudo-random graphs.

2. Quadrupartions

Problem poser: Imre Bárány (Alfréd Rényi Institute of Mathematics, Budapest, Hungary, e-mail: barany@renyi.hu)
Co-author: Branko Grünbaum (Department of Mathematics, University of Washington)

A few decades ago, Branko Grünbaum raised the following problem which has never appeared in print:

Conjecture 1. *If K is a convex set in the plane whose area is one, then for each $t \in [0, 1/4]$ there exist two orthogonal lines that partition K in four pieces of areas t, t, $(1/2 - t)$, $(1/2 - t)$ in clockwise order.*

An elementary continuity (or topological) argument shows that the conjecture is true when $t = 0$ and when $t = 1/4$. It seems that topology, or topology alone, is not enough for the general case.

Here is an even stronger version of the conjecture. Assume that a nice probability measure is given in the plane. (Here "nice" means, say, that it is a Borel measure which is zero on every line segment.) Again, let $t \in [0, 1/4]$. Is it true, then, that there exists a pair of orthogonal lines partitioning the plane into four pieces with measures t, t, $1/2 - t$, $1/2 - t$ in clockwise order? (Again, this is true for $t = 0$ and $t = 1/4$.) We venture to conjecture that the answer in general is "no":

Conjecture 2. *There is a nice probability measure in the plane, so that, for some $t \in (0, 1/4)$, no pair of orthogonal lines partitions the plane into four pieces of measure t, t, $1/2 - t$, $1/2 - t$ in clockwise order.*

3. Chromatic Number of a Random Subgraph

Problem poser: Boris Bukh (Princeton University / UCLA, e-mail: B.Bukh@dpmms.cam.uc.uk)

Chromatic number of a graph G is denoted $\chi(G)$. Let G_p be the random subgraph of G obtained from G by keeping each edge with probability p and removing it with probability $1 - p$ independently at random. For example,

$G_{1/2}$ is the subgraph chosen uniformly at random among all subgraphs of G. It is easy to see that

$$\mathbb{E}\big[\chi(G_p)\big] \geq \chi(G)^p$$

whenever $1/p$ is a positive integer. Is there an $\varepsilon > 0$ such that

$$\mathbb{E}\big[\chi(G_{1/2-1/1000})\big] \geq \chi(G)^{1/3+\varepsilon}?$$

Perhaps it is even true that

$$\mathbb{E}\big[\chi(G_p)\big] \geq c(p)\frac{\chi(G)}{\log \chi(G)}$$

for every fixed $0 < p < 1$, and a positive constant $c(p)$ depending only on p. It is known that the inequality

$$\mathbb{E}\big[\chi(G_p)\big] \geq c(p)\frac{\chi(G)}{\log |V(G)|}$$

holds.

4. Two Conjectures on Quasi-Kernels

Problem poser: Péter L. Erdős (Alfréd Rényi Institute of Mathematics, Budapest, Hungary, e-mail: elp@renyi.hu)

Let $D = (V, \vec{E})$ be (finite or infinite) directed graph. An independent vertex subset $A \subset V$ is a *quasi-kernel* (also known as *semi-kernel*) iff for each point v there is a path of length at most 2 from some point of A to v. Similarly, the independent vertex subset $B \subset V$ is a *quasi-sink,* iff for each point v there is a path of length at most 2 from v to some point of A.

It is a well-known fact, that every finite (table-tennis) tournament has a (single-point) quasi-kernel (quasi-sink). In 1973 the following nice generalization was proved:

Theorem 1 (V. Chvátal – L. Lovász [5]). *Every finite directed graph contains a quasi-kernel (quasi-sink).*

In 1976 the following conjecture was stated:

Conjecture 2 (P. L. Erdős – L. A. Székely). *Assume that in the finite digraph $D = (V, \vec{E})$ for each vertex $v \in V$ the* indegree $d^-(v) \geq 1$. *Then there exists a quasi-kernel A with the property $|A| \leq |V|/2$. For example the disjoint union of oriented C_4's satisfies this with equality.*

Theorem 1 does not hold in case of infinite directed graphs: for example if Z denotes the directed graph of all integers where each edge is directed "upwards", clearly there is nor quasi-kernel neither quasi-sink. However its vertex set can be easily partitioned into two subsets, such that one spanned sub-tournament contains a quasi-kernel, while the other one contains a quasi-sink.

Conjecture 3 (P. L. Erdős – L. Soukup (2008) [8]). *Every (countable) infinite directed graph D can be partitioned into two vertex classes, such that one spanned subgraph contains a quasi-kernel, while the other one contains a quasi-sink.*

5. Partitioning Graphs into Large Stars

Problem poser: Jan Kratochvil (Department of Applied Mathematics, Charles University, Prague, Czech Republic,
e-mail: honza@kam.ms.mff.cuni.cz)

Informally, the question is to identify sufficient conditions that would guarantee that the vertex set of a graph under consideration can be partitioned into large enough stars. In particular, we ask whether large minimum degree or large vertex connectedness are such sufficient conditions.

Formally, for a graph $G = (V, E)$, we define $\mathrm{sc}(G)$ to be the maximum s such that G can be partitioned into disjoint stars with at least s rays each (i.e. $V = \bigcup_{i=1}^{t} V_i$, such that for each i, $|V_i| \geq s + 1$ and $G[V_i]$ contains a vertex adjacent to all other its vertices). We further define two functions

$$\mathrm{sc}_\delta(k) = \min_{G:\,\delta(G)\geq k} \mathrm{sc}(G), \quad \text{and}$$

$$\mathrm{sc}_\kappa(k) = \min_{G\ k\text{-vertex-connected}} \mathrm{sc}(G).$$

Problems. 1) Is $\lim_{k\to\infty} \mathrm{sc}_\delta(k) = \infty$?

2) Is $\lim_{k\to\infty} \mathrm{sc}_\kappa(k) = \infty$?

3) If yes, determine, as tight as possible, the growth rates of these functions.

Comments. The motivation for the problem comes from TCS, namely exact exponential time algorithms for the so called $L(2,1)$-labeling problem (joint work with F. Havet, M. Klazar, D. Kratsch, and M. Liedloff). The problem has been solved by N. Alon and N. Wormald; see their article in the present volume.

6. Open Problems on Girth

Problem poser: Nathan Linial (School of Computer Science and Engineering, Hebrew University, Jerusalem, Israel, e-mail: nati@cs.huji.ac.il)

We recall that the *girth* of a graph G is the length of the shortest cycle in G. Let us denote by $g(d,n)$ the highest girth of a d-regular n-vertex graph. Our main interest is is the case where $d \geq 3$ is fixed and n is large. The current best asymptotic bounds on this function are

$$\big(2+o(1)\big)\frac{\log n}{\log(d-1)} \geq g(d,n) \geq \left(\frac{4}{3}-o(1)\right)\frac{\log n}{\log(d-1)}.$$

Both sides of this inequality raise interesting questions. The lower bound comes from a famous paper by Lubotzky Phillips and Sarnak and is based on deep results from representation theory. Using combinatorial and probabilistic methods it is easy to prove the lower bound with coefficient 1 instead of $\frac{4}{3}$. Any new proof using combinatorial and probabilistic proof that yields $1+\varepsilon_0$ for some ε_0 that is bounded away from zero would be of interest.

The upper bound (also called the Moore Bound) is quite obvious. I conjecture, though, that there is some $\delta > 0$ such that for every integer $d \geq 3$ and large enough n, there holds:

$$(2-\delta)\frac{\log n}{\log(d-1)} \geq g(d,n).$$

7. Chromatic Number of Subgraph

Problem poser: László Lovász (Eötvös Loránd University, Budapest, Hungary, e-mail: lovasz@cs.elte.hu)

An old problem: Is it true that every simple graph without isolated nodes that is not complete contains two adjacent nodes u and v such that $\chi(G-u-v) \geq \chi(G)-1$?

8. Edmonds Graphs

Problem poser: Claudio L. Lucchesi (University of Campinas, Sao Paulo, Brazil, e-mail: lucchesi@ic.unicamp.br)
Co-authors: Marcelo H. de Carvalho (UFMS – Brazil), U. S. R. Murty (University of Waterloo – Canada)

Edmonds 1965. A vector $\mathbf{x}$ in $\Re^E$ belongs to the perfect matching polytope of a graph G iff:

$$\begin{aligned} \mathbf{x}(e) &\geq 0 \quad \forall e \in E && \text{(non-negativity)} \\ \mathbf{x}\big(\partial(v)\big) &= 1 \quad \forall v \in V && \text{(degree constraints)} \\ \mathbf{x}\big(\partial(S)\big) &\geq 1 \quad \forall S \subset V,\ |S| \text{ odd} && \text{(odd set constraints)} \end{aligned}$$

Bipartite graphs do not need odd set constraints.

An *Edmonds graph* is one for which the odd set constraints are needed.

Question. Is the class of Edmonds graphs in co-$\mathcal{NP}$?

We know that the class is in $\mathcal{NP}$.

Let G be a graph. Graph G is *matching covered* if it is non-trivial, connected and every edge lies in a perfect matching. For a subset S of V, the *cut* $\partial(S)$ consists of those edges having precisely one end in S. A cut $\partial(S)$ is *tight* if every perfect matching of G has precisely one edge in $\partial(S)$, and it is *separating* if both G/S and $G/\overline{S}$ are also matching covered. Every tight cut is separating, but not every separating cut is tight. In a bipartite graph, every separating cut is tight.

Carvalho, Lucchesi, Murty (JCT-B, 2004). A brick G is Edmonds iff it has a separating cut that is not tight.

Reed, Wakabayashi 2004. A brick is Edmonds iff it has two disjoint odd cycles C_1 and C_2 such that $G - V(C_1) - V(C_2)$ has a perfect matching.

Corollary. *If a graph is Edmonds then it must have two disjoint odd cycles.*

Example. Odd wheels and Möbius ladders are not Edmonds.

We have shown [Discrete Math., 2006] that, essentially, the only planar matching covered graphs that are not Edmonds are the bipartite graphs and the odd wheels.

9. Counting Convex Colorings of Graphs

Problem poser: Johann A. Makowsky (Dept. of Computer Science, Technion Haifa, Israel, e-mail: janos@cs.technion.ac.il)

A vertex coloring of a graph $G = (V, E)$ with k colors ($k \in \mathbb{N}$) is a function $f : V \to [k]$. f is *convex* if for every $i \in [k]$ the colorclass $f^{-1}(i)$ induces a connected subgraph. For a partial function $f_0 : V \to [k]$ we say that f_0 is convex if there a is a total function f extending f_0 which is convex. In this case we also say that f is a *convex extension* of f_0. Convex extensions of partial colorings of trees have been introduced in the context of phylogenetic trees by S. Moran and S. Snir [15].

The existence problem of convex colorings for an arbitrary graph G is easily solved by trying to color every connected component by one color, and only depends on the number of colors available and the number of connected components of G. It follows from [13, 10] that the number of convex colorings of a graph G is a polynomial in k, which we denote by $\mathrm{conv}\,(G, k)$. For $k = 1$ we have $\mathrm{conv}\,(G, 1) = 1$, if G is connected, and $\mathrm{conv}\,(G, 1) = 0$ otherwise.

Problem. What is the complexity of computing $\mathrm{conv}\,(G, 2)$?

We think that this problem is $\sharp P$-complete. Let E_n be the empty graph on n vertices. One easily verifies that

$$\mathrm{conv}\,(G \sqcup E_n, k) = \mathrm{conv}\,(G, k-n) \cdot \big(k \cdot (k-1) \cdot \ldots \cdot (k-n+1)\big)$$

where $\sqcup$ denotes the disjoint union of graphs. This shows, similar as in [11], that, if evaluating $\mathrm{conv}\,(G, 2)$ is $\sharp P$-hard, so is evaluating $\mathrm{conv}\,(G, k)$ for $k \geq 2$.

Using methods from [7] one can prove that for fixed k, evaluating conv (G, k) is P-time computable on graphs of tree-width at most t.

Comments. The problem has been solved by A. J. Goodall an S. D. Noble, see [9].

10. Maximum Number of $K_{r,r}$ Subgraphs

Problem poser: Dániel Marx (Budapest University of Technology and Economics, Budapest, Hungary, e-mail: dmarx@cs.bme.hu)

Let G be an n-vertex graph that has no $K_{r,r+1}$ subgraph. What is the maximum number of $K_{r,r}$ subgraphs in the graph? In particular, can we give an upper bound of the form $O(n^c)$ (for some absolute constant c) or of the form $f(r) \cdot n^c$ (for some function f and absolute constant c)? Is it possible to construct graphs where this number is $n^{\Omega(r)}$?

11. On Hajnal's Triangle-Free Game

Problem poser: Ákos Seress (The Ohio State University, e-mail: akos@math.ohio-state.edu)

András Hajnal proposed the following graph game in the early 1990's. Beginning with the empty graph on n vertices, two players, A and B, alternatingly draw edges. We do not distinguish the edges drawn by A and B. The loser is the player who cannot make a move without completing a triangle.

The winner of the game is known only for $n \leq 15$ [3, 17, 19] and the main problem is to determine the winner for all n.

Some results were obtained in variants of the game, when we place restrictions on the graph created by the players. If the non-isolated vertices of the graph drawn must always be in one connected component then the winner depends only on the parity of n [19]. In another variant, we require that the graph created by the players has maximal degree at most k. This version is surprisingly difficult even for $k = 3$, and the winner is known only for $k = 2$ and $k = 3$ [14]. After some erratic behaviour for small n, the

winner depends on the value n mod 2 and n mod 4 in the cases $k = 2$ and 3, respectively. The restricted degree variant makes sense even if we drop the triangle-free requirement, and also in this version the winner is known only for $k = 2$ and $k = 3$ [14].

12. Is Two Coloring Always Polynomial?

PROBLEM POSER: Joel Spencer (Department of Computer Science and Department of Mathematics, Courant Institute, New York, USA, e-mail: spencer@cims.nyu.edu)

Suppose A_α is a family of k-element subsets of an n-set Ω with each set intersecting at most d other sets. The Lovász Local Lemma says that if $4d2^{1-k} < 1$ (this can be slightly strengthened, for the best bound, see [18]) there necessarily *exists* a two coloring of Ω with no A_α monochromatic. Is there, for every such fixed d, k a polynomial (in n) time algorithm that *finds* that coloring? Beck has shown that the answer is yes for a wide class of d, k. We conjecture the general answer is no. In somewhat imprecise complexity terms: that there are fixed d, k for which the problem of finding a coloring can be placed in a class of problems which are generally regarded as not solvable in polynomial time.

Added in Proof. Such a randomized polynomial time algorithm has been found recently, see [16]. The existence of a deterministic polynomial time algorithm is still open in this generality, but such an algorithm is given in case the slightly stronger condition $d = O\left(2^{(1-\varepsilon)k}\right)$ holds, see [4].

13. Pizza Problem

PROBLEM POSER: Peter Winkler (Department of Mathematics, Dartmouth, Hanover, New Hampshire, USA, e-mail: peter.winkler@dartmouth.edu)

Alice and Bob share a pizza; the pizza is round and sliced by radial cuts into some arbitrary number of pieces of various sizes. Alice chooses any slice to eat, after which Bob and Alice alternate taking slices, but must always take a slice which is adjacent to some previously-taken slice.

Conjecture: Alice can always get at least 4/9 of the pizza.

Note: 4/9 is best possible; 1/3 is easy.

Comment: based on a puzzle by Dan Brown which asked whether Alice can guarantee 1/2.

Comments. The problem has been solved by J. Cibulka, J. Kynčl, V. Mészáros, R. Stolař and P. Valtr; see their article in the present volume. The problem also has been solved simultaneously and independently by the group consisting of K. Knaurer, T. Uekerdt and P. Micek.

References

[1] N. Alon and E. Berger, The Grothendieck constant of random and pseudo-random graphs, *Discrete Optim.*, **5** (2008), no. 2, 323–327.

[2] N. Alon, K. Makarychev, Y. Makarychev and A. Naor, Quadratic forms on graphs, *Invent. Math.*, **163** (2006), no. 3, 499–522.

[3] S. C. Cater, F. Harary and R. W. Robinson, *One-color triangle avoidance games,* Proceedings of the Thirty-second Southeastern International Conference on Combinatorics, Graph Theory and Computing (Baton Rouge, LA, 2001), vol. 153, 2001, pp. 211–221.

[4] K. Chandrasekaran, N. Goyal and B. Haeupler, Determininistic algorithms for the Lovász Local Lemma, in: *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithm* (SODA 2010).

[5] V. Chvátal and L. Lovász, *Every directed graph has a semi-kernel,* Hypergraph Seminar (Proc. First Working Sem., Ohio State Univ., Columbus, Ohio, 1972; dedicated to Arnold Ross), Springer, Berlin, 1974, pp. 175. Lecture Notes in Math., Vol. 411.

[6] J. Cibulka, J. Kynčl, V. Mészáros, R. Stolař and P. Valtr, *Solution of Peter Winkler's pizza problem,* this volume.

[7] B. Courcelle, J. A. Makowsky and U. Rotics, On the fixed parameter complexity of graph enumeration problems definable in monadic second-order logic, *Discrete Appl. Math.*, **108** (2001), no. 1-2, 23–52, International Workshop on Graph-Theoretic Concepts in Computer Science (Smolenice Castle, 1998).

[8] P. L. Erdős and L. Soukup, Quasi-kernels and quasi-sinks in infinite graphs, *Discrete Mathematics,* **309** (2009), no. 10, 3040–3048.

[9] A. J. Goodall and S. D. Noble, Counting cocircuits and convex two-colourings is #P-complete, preprint, arXiv: 0810.2042, 2008. `http://arxiv.org/pdf/0810.2042`

[10] T. Kotek, J. A. Makowsky and B. Zilber, *On counting generalized colorings,* CSL '08: Proceedings of the 22nd international workshop on Computer Science Logic (Berlin, Heidelberg), Springer-Verlag, 2008, pp. 339–353.

[11] N. Linial, Hard enumeration problems in geometry and combinatorics, *SIAM J. Algebraic Discrete Methods*, **7** (1986), no. 2, 331–335.

[12] L. Lovász, On the Shannon capacity of a graph, *IEEE Trans. Inform. Theory*, **25** (1979), no. 1, 1–7.

[13] J. A. Makowsky and B. Zilber, *Polynomial invariants of graphs and totally categorical theories*, MODNET Preprint No. 21, `http://www.logique.jussieu.fr/modnet/Publications/Preprint%20 server`, 2006.

[14] N. Mehta and Á. Seress, *Graph Games*, manuscript, in preparation.

[15] S. Moran and S. Snir, Efficient approximation of convex recolorings, *J. Comput. System Sci.*, **73** (2007), no. 7, 1078–1089.

[16] R. Moser and G. Tardos, A constructive proof of the general Lovász Local Lemma, *Journal of the ACM*, **57** (2010) (2), Art. 11.

[17] P. Pralat, A note on the one-colour avoidance game on graphs, *Journal of Combinatorial Mathematics and Combinatorial Computing*, 9pp (programs in C/C++), to appear.

[18] J. Radhakrishnan and A. Srinivasan, Improved bounds and algorithms for hypergraps 2-coloring, *Random Structures Algorithms*, **16(3)** (2000).

[19] Á. Seress, On Hajnal's triangle-free game, *Graphs Combin.*, **8** (1992), no. 1, 75–79.